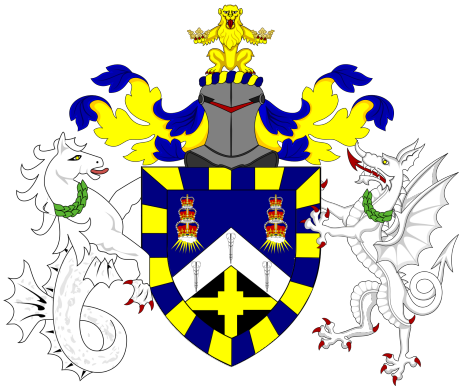
Processing of Cargo Proteins By The Type IX Secretion System



Benjamin James Dorgan

School of Biological And Chemical Sciences

Queen Mary University of London

This dissertation is submitted for the degree of

Doctor of Philosophy

2020

Declaration

I, Benjamin James Dorgan, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below. I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material. I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis. I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university. The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Benjamin James Dorgan

2020

Acknowledgements

Firstly I would like to thank Dr James Garnett for giving me the opportunity to undertake my PhD in his lab and for his supervision for its duration. His guidance and support has allowed me to develop my technical lab skills and my ability to be an independent researcher. I would also like to thank Dr Vidya Darbari for coming into the project to offer her support for this PhD. The supervision of Prof Mike Curtis has been invaluable and his critical appraisal of the research has helped me to see my own work in a new way. I would also like to thank Dr Mark Roberts for his support during this project and for always being free to discuss new ideas and directions.

I would also like to thank everyone in the Garnett lab for assistance with this project. Theo Portlock has followed the same journey as me and has been a constant friend with which to discuss research ideas and the best brands of packet noodles. I would like to thank Sunjun Wang for sharing in my difficulties in working with the T9SS, for helping in the preparation of protein samples and his permanent smile despite the situation. I would also like to thank Dr Paula Corsini, Dr Lee Sewell and Dr Saima Rehman for giving me knowledgable advice and suggestions during the course of this project. I would like to give thanks to Dr Joe Aduse-Opoku, from Mike Curtis' lab, for his support and for sharing his expertise of microbiology and *P. gingivalis*. Joe has always been ready to help and give assistance in any way possible.

I would like to thank the MRC Biomedical NMR Centre, particularly Dr Alain Oregioni and Dr Geoff Kelly, for collection of NMR spectra. I would also like to thank

them for their educational support regarding NMR spectroscopy. I would also like to give thanks to Dr Chris Lorenz for helping me to learn and set up molecular dynamics simulation.

I would like to thank the LiDo programme and its organisers for giving me the opportunity to carry out this PhD and for all the additional activities to support my professional development. I would also like to thank the Cabinet Office Policy Lab for the opportunity to do my PIPs as part of their great team.

I would also like to thank my parents for encouraging me to pursue and achieve my aspirations. They have always been there to support me when I need it most.

Lastly I would like to thank the most important person in my life. Abbie has been with me from the start and has always given me her love and support, particularly when things are most difficult. Without her I don't know how I would have made it through.

Abstract

Porphyromonas gingivalis is a Gram-negative oral anaerobic pathogen and is one of the key causative agents of the periodontitis. *P. gingivalis* utilises a range of virulence factors to drive pathogenesis that are exported and attached to the cell surface via the Type 9 Secretion System (T9SS). All cargo proteins possess a conserved C-terminal signal domain (CTD) which is recognised by the T9SS via an unknown mechanism. The T9SS transports cargo proteins to the outer membrane where the CTD is remove and the cargo is anchored to the cell surface. The aim of this thesis was to provide structural characterisation of the CTD and to understand how it interacts with components of the T9SS. To achieve this, solution state NMR was used to study the structure and dynamics of the CTD from RgpB, showing it to be a seven stranded β sandwich, in alignment with the crystallographic model. The model presented here, supported by NMR relaxation data, suggests RgpB- CTD a monomer in solution, unlike the crystal structure, in which a dimeric state is observed. It has previously been shown that the outer-membrane β-barrel protein PorV is able to interact with T9SS cargo proteins. It has been possible to recombinantly express and purify the full length PorV protein and work is ongoing to demonstrate that it interacts with RpgB-CTD via a pulldown assay. In this study a predicted structure of *P. gingivalis* PorV was generated and simulated in a lipid bilayer using GROMACs. Representative conformations were used to generate a structural ensemble for molecular docking with the NMR model of RgpB-CTD. Binding was examined between the two proteins and

multiple binding poses were observed. From these models two possible mechanisms of CTD recognition by PorV are suggested. Future work to experimentally study this interaction will allow for the models to refined further.

Contents

List of Figures

List of Tables

Nomenclature

A-LPS	Anionic lipopolysaccharide
APS	Anionic polysaccharide
AU	Approximately unbiased
Bo	Externally Applied Static Magnetic Field
B_1	Applied Magnetic Field
BHI	Brain heart infusion
BHI-H	Brain heart infusion with haemin
B_{Local}	Induced Magnetic Field
BN-PAGE	Blue native polyacrylamide gel electrophoresis
COSY	Correlation spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill
CRD	Carboxypeptidase regulatory domain-like fold
Cryo-EM	Cryo-electron microscopy
Cryo-ET	Cryo-electron tomography
CTD	C-terminal domain
DDM	n-dodecyl-{b}-D-maltopyranoside
DMSO	Dimethyl sulfoxide
DSP	Dithiobis succinimidyl propionate

DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
ECM	Extracellular matrix
EDSL	Electron-dense surface layer
FID	Free Induction Decay
γ	Gyromagnetic Ratio
GRAVY	Grand average of hydropathy
HA	haemagglutinin/adhesin
HSQC	Heteronuclear signle quantum coherence
I	Spin Quantum Number
IPTG	Isopropyl \{}beta-D-1-thiogalactopyranoside
Kgp	Lys gingipain
LB	Luria broth
LDAO	Lauryldimethylamine oxide
LPS	Lipopolysaccharide
μ	Magnetic Moment
m	Magnetic Quantum Number
M	Bulk Magnetisation
MD	Molecular dynamics
MMP	metalloproteinases
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NTA	nitrilotriacetic acid
NTP	N-terminal pro-domain

ω_0	Larmor Frequency
O-LPS	Conventional lipopolysaccharide
OG	octyl-{b}-D-glucoside
OMV	Outer membrane vehsicle
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPE	1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
PorSS	Por secretion system
PPAD	peptidylarginine deiminase
PPI	Protein-protein interaction
RF	Radio Frequency
RgpA	Arg gingipain A
RgpB	Arg gingipain B
RMSD	Root mean square deviation
RMSD	Root mean square deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SMA2000	styrene maleic anhydride
T_1	Longitudinal Relaxation
T_2	Transverse Relaxation
T9SS	Type 9 secretion system

TLCK	Nα-Tosyl-L-lysine chloromethyl ketone hydrochloride
TLR	Toll-like receptor
TOCSY	Total correlation spectroscopy
TPR	Tetratricopeptide repeat

Chapter 1

Introduction

1.1 Periodontal Disease

Periodontal disease is an inflammatory condition affecting the gingiva and the supporting structures of the periodontium [1]. In severe cases it can lead to the destruction of the periodontal ligament, and alveolar bone leading to tooth loss [2]. Periodontal disease is classified into two broad groups: gingivitis and periodontitis - both of which can be either local or general [3]. Gingivitis is classified as inflammation of the gingival tissue and accumulation of dental plaque but with no damage to the gingiva and periodontium (Figure 1.1.1B). Gingivitis is a reversible state but it can progress to periodontitis. Periodontitis is classified as irreversible inflammation of the periodontal tissue with damage caused to the periodontal ligament, alveolar bone and tooth (Figure 1.1.1C). This causes the formation of the periodontal pocket - a cavity formed between the gum and the tooth that provides a suitable environment for further colonisation by the oral biofilm [4]. Chronic periodontitis is widely prevalent with around 10% of the global population being affected by severe periodontitis [5]. It is particularly prevalent in the western world with an estimated 46% of the US adult population affected [6]. Many risk factors that contribute to the development of chronic

periodontitis some of which can be modified and controlled, some which cannot [7]. Modifiable factors include smoking, obesity, diet, stress, oral hygiene and the composition of the oral microbiome. Non-modifiable factors include genetic polymorphisms and underlying systemic conditions such as AIDS [8] and Downs Syndrome [9]. The development of chronic periodontitis is in turn associated with increased susceptibility to other diseases including cardiovascular disease [10], diabetes [11], arthritis [12] and Alzheimer's disease [13].

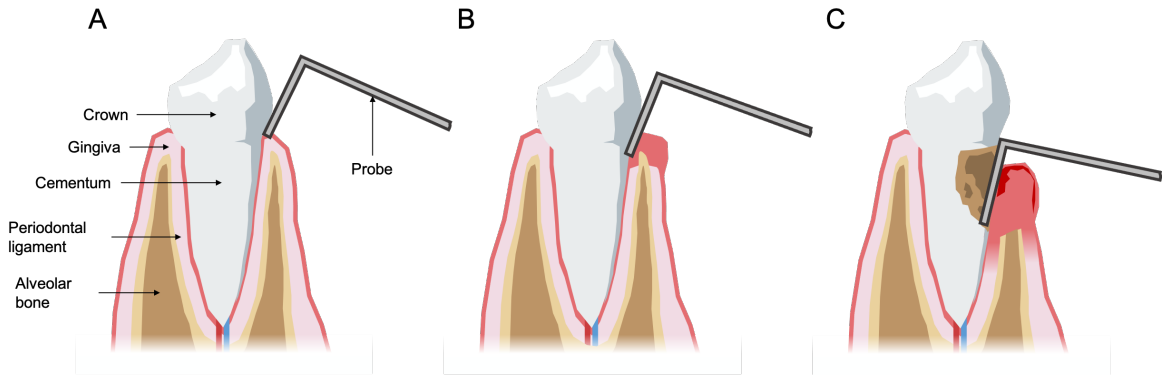


Figure 1.1.1: Progression of periodontal disease. (A) Healthy gingiva with no inflammation and pocket depth of less than 3 mm. (B) Tooth with gingivitis, characterised by a build up of plaque, swollen gums and a deepening of the periodontal pocket. (C) Tooth with periodontitis, characterised by severe inflammation, destruction of the alveolar bone and recession of the gum leading to the formation of a periodontal pocket with a depth greater than 5 mm. Advanced stages of periodontitis can lead to loosening or loss of the tooth. (Adapted from [14]).

1.2 The Role of *P. gingivalis* in Periodontal Disease

The development of periodontal disease is not caused by a single pathogenic organism but by a consortium of pathogenic microbes. As the biofilm develops the population of Gram-negative anaerobes, including the potential pathogens *Porphyromonas gingivalis, Treponema denticola* and *Tannerella forsythia*, increases [15]. These three

bacterial species are known as the 'red complex' and often co-associate during disease progression [16]. The presence of these potentially pathogenic species does not necessitate the development of a pathological state, but can swing the balance between a healthy oral biofilm to a pro-inflammatory one [17]. These pro-inflammatory biofilms are termed 'dysbiotic', meaning that there is an imbalance in the relative microbial composition and/or influence of the biofilm leading to inflammatory disease [18].

Of the microbes associated with the development of periodontal disease, *P. gingivalis* appears to play a particularly central role and was one of the earliest pathogens found to associate with disease progression [19]. *P. gingivalis* is a Gram-negative, non-motile, asaccharolytic, obligate anaerobe that can invade the oral epithelium and periodontal tissue. *P. gingivalis* primarily inhabits the subgingival sulcus as part of a complex biofilm. Because *P. gingivalis* is an obligate anaerobe it is a secondary coloniser that adheres to primary colonisers such as *Corynebacterium* and *Streptococcus* [20]. Due to the low availability of sugars in the periodontal pocket, *P. gingivalis* relies on the fermentation of dipeptides as a source of energy [21–23].

In recent years a hypothesis has developed that *P. gingivalis* drives disease progression by interfering with and impairing host immunity to favour the development of a dysbiotic biofilm [17]. The 'keystone pathogen hypothesis' is supported by evidence from studies in mice that even very low colonisation levels, *P. gingivalis* can significantly alter the composition of the periodontal biofilm and induce periodontitis [24]. Furthermore, colonisation of germ-free mice with *P. gingivalis* alone did not lead to the development of periodontitis demonstrating the necessity of commensal microbes in pathogenesis. This hypothesis is consistent with observations that there is a low abundance of *P. gingivalis* in periodontitis patients [25–27], but a very high prevalence and association with periodontitis pathogenesis [28, 29].

1.3 *P. gingivalis* Virulence Factors

Due to the crucial role of *P. gingivalis* in periodontitis, there has been a considerable research effort focused on understanding the virulence factors produced by *P. gingivalis* that can impair host immunity, increase inflammation and alter the composition of the oral microbiota [1]. *P. gingivalis* is known to produce a range of virulence factors that enable invasion, inflammation, evasion of innate immunity and nutrient acquisition. A brief description of the main virulence factors produced by *P. gingivalis* is given here.

Multiple strains of *P. gingivalis* are known to produce a capsule that surrounds the cell and shields surface components from the host. Encapsulated strains of *P. gingivalis* are more virulent than non-encapsulated strains when studied using a mouse infection model, with non-encapsulated strains forming non-invasive, localised abscesses [30, 31]. Encapsulated strains are associated with a reduced pro-inflammatory response, phagocytosis and cytokine and chemokine production [32, 33]. This enables encapsulated strains to evade host immunity and increase bacterial survival.

P. gingivalis fimbriae have been shown to mediate adhesion to other bacterial species, oral epithelial cells and tooth surfaces with fimbriae deficient mutants showing a reduced ability to attach and invade oral epithelial cells [19, 34]. *P. gingivalis* expresses two types of fimbriae, the first consisting of the FimA/fimbrillin subunit to form the major fimbriae and the second consisting of the Mfa subunit to form the minor fimbriae. While FimA appears to function primarily as an adhesin, the role of the Mfa minor fimbriae is less clear [1]. It has been demonstrated that the minor fimbriae can stimulate the production of IL-1α, IL-1β, IL-6 and TNF-α by macrophages [35]. This increase in pro-inflammatory signals has been observed in rat infection models and is suggested to promote resorption of the alveolar bone [36, 37].

Lipopolysaccharide (LPS) is an important component of the outer membrane of Gram-negative bacterial. It is comprised of Lipid A, a core oligosaccharide and a distal

O-antigen of repetitive glycan polymers. There are two major forms of LPS molecule produced by *P. gingivalis*, the conventional O-antigen polysaccharide linked LPS (O-LPS) and anionic-LPS (A-LPS) that possesses an anionic polysaccharide (APS) [38]. There is also variation in the acylation profile of Lipid A, with tri-, tetra- and penta-acyl forms being observed [39]. The Lipid A component of LPS has been shown to interfere with innate immunity by interacting with toll-like receptor (TLR)-2 and -4 [40]. It has been demonstrated that *P. gingivalis* LPS is a less potent activator of pro-inflammatory signals than *Escherichia coli* LPS, which may also help *P. gingivalis* to evade innate immunity [41]. *P. gingivalis* LPS is also capable of inhibiting osteoblast differentiation of periodontal ligament stem cells which play an important role in the regeneration of periodontal tissue [42].

P. gingivalis secretes a range of proteolytic, hydrolytic and lipolytic enzymes that increase virulence and drive disease progression. Many of these enzymes are secreted from the cell and they often remain attached to the surface. Of the different enzymes classes produced by *P. gingivalis,* it is the proteolytic enzymes that appear to correlate most closely with pathogenesis [43]. *P. gingivalis* mostly produces cysteine proteases or serine proteases. The cysteine proteases, commonly called gingipains, are involved in multiple detrimental interactions with the host and activate other surface-bound components, such as the fimbriae [14].

P. gingivalis forms outer membrane vesicles (OMVs) that are highly enriched in LPS and membrane-associated virulence factors, particularly gingipains [44]. The vesicles are spherical structures, with a diameter of ~50-300 nm, composed of a single lipid bilayer formed via the blebbing of the bacterial outer membrane [45]. OMVs contribute to the destruction of host tissue by acting as an effective vehicle for the delivery of proteases and other virulence factors [46]. For example, *P. gingivalis* OMVs are capable of entering epithelial cells and interfering with host signalling pathways

to prevent cell migration [47]. It is also believed that the release of OMVs contributes to more effective colonisation and evasion of host immune defences via the activity of surface-associated gingipains [45].

1.4 The Role of Gingipains in Periodontal Pathology

Of the virulence factors produced by *P. gingivalis*, there has been a particular focus on understanding a group of extracellular proteases, discovered in the early '90s, that closely associate with virulence and have the ability to degrade a broad spectrum of host proteins [48–51]. The gingipains, as they are known, are cysteine proteases that cleave at the C-terminus of arginine (gingipain R; or RgpA and RgpB) or lysine (gingipain K; or Kgp) residues. Collectively, the gingipains account for ~85% of *P. gingivalis* extracellular proteolytic activity at the site of infection [52]. The proteolytic activity of the gingipains is involved in many pathogenetic processes including the degradation of extracellular matrix proteins, complement factors, immunoglobulins, cell surface receptors and deregulation of host signalling pathways.

One way in which the gingipains assist in the survival of *P. gingivalis* is by increasing the availability of haemin required for growth. RgpA and RgpB can activate prothrombin to thrombin by cleavage of coagulation factors IX and X into an active state [53, 54]. Gingipains are also capable of degrading fibrinogen and fibrin, further deregulating the coagulation cascade, leading to increased bleeding and inflammation at the site of infection [55, 56]. Furthermore, the gingipains can cleave host haem-binding proteins to release a usable source of haem and short peptides necessary for *P. gingivalis* growth [**?**]. This has been suggested to contribute to the extensive proliferation in periodontal pockets where there is a high abundance of erythrocytes.

Gingipains cause damage to the host extracellular matrix (ECM) via the proteolytic activation of matrix metalloproteinases (MMPs) secreted by many host cells as inactive pro-MMP precursors [57]. The activation of these enzymes causes them to degrade collagen fibres, leading to the destruction of ECM and the periodontal ligament. The destructive activity of activated MMPs contributes to the formation of a deep periodontal pocket [58].

It has been observed that strains of *P. gingivalis* that produce gingipains can significantly decrease the bactericidal activity of human serum. It was found that low concentrations of gingipains could activate complement factors C3, C4 and C5, perhaps to increase local inflammation and assisting in colonisation [59]. However, at high concentrations the gingipains were found to degrade C3, C4 and C5 to prevent activation of the complement cascade accounting for the bactericidal activity of human serum. RgpA and B can further interfere with the complement pathway by activating complement factor C5 to create subversive cross-talk between the C5a receptor and TLR-2, preventing phagocytosis [60].

IL-8 is a pro-inflammatory chemokine that is produced by a variety of cell types including lymphocytes, monocytes, neutrophils, endothelial cells and epithelial cells [61]. IL-8 attracts and neutrophils towards sites of inflammation and is also a potent activator of neutrophil activity [62]. Soluble gingipains that are released from the cell cleave IL-8 into a shorter and more potent form, whereas surface-bound gingipains degrade IL-8, eliminating its chemotactic function [63]. This dual process creates a chemotactic gradient away from the site of infection helping *P. gingivalis* to evade the host immune system. Gingipains can also attack host receptors directly. One target is CD-14, the major receptor of LPS expressed by monocytes [64] and gingival fibroblasts [65]. By degrading this receptor, renders the host less able to detect LPS reducing the production of pro-inflammatory signals.

Gingipain activity is also necessary for the processing and maturation of surface-bound proteins expressed by *P. gingivalis*. *P. gingivalis* produces a range of haemagglutinin/adhesin (HA) domains that are encoded by *rgpA, kgp* and *hagA*. The gingipain encoded adhesin domains are autolytically processed and form a non-covalent gingipain-adhesin complex that is associated with the OM [66]. The HagA encoded HA domains are also proteolytically processed by the gingipains [67]. The adhesin domains are necessary for the haemagglutination and haemolysis of erythrocytes and binding to haemoglobin to facilitate the acquisition of haem by *P. gingivalis* [68–70]. The maturation of both the major and minor fimbriae is also dependent on proteolytic processing by RgpA and B [71, 72].

1.5 Gingipains Possess a Conserved C-terminal Domain

Ever since the discovery of the gingipains, it has been reasoned that there must a mechanism by which these proteins are secreted to the cell surface [73]. The gingipains possess an N-terminal Sec signalling peptide that mediates transport of the protein into the periplasm where the protein folds [74]. An N-terminal pro-domain (NTP) functions as a chaperone and maintains the protease in an inactive state to prevent cleavage of cellular proteins [75]. In the mature surface-bound RgpB, the NTP is removed, along with 72 amino acids at the C-terminus. RgpA and Kgp also possess a C-terminal HA region that undergoes autolytic processing and remains non-covalently linked to the catalytic domain (Figure 1.5.1) [66, 67]. The HA region of RgpA consists of four domains referred to as RgpA1-4 [66]. The number of domains comprising the Kgp HA region varies between sequenced strains of *P. gingivalis* [14]. The W50 strain, used for experimental work in this thesis, contains five HA domains referred to as

KgpA1-5 [67].

BN-PAGE analysis of surface associated complexes from *P. gingivlalis* has shown that gingipain-adhesin complexes of Kgp and RgpA form either homodimeric or heterodimeric (Kgp:RgpA) complexes [76]. These complexes are found predominantly bound to the cell-surface with very little protein released to the extracellular milieu [74].

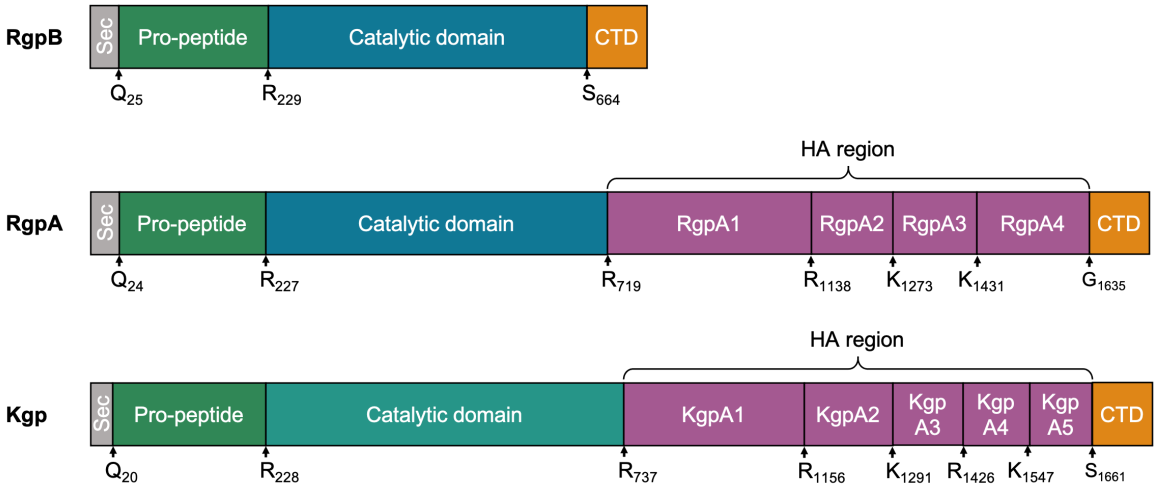


Figure 1.5.1: Schematic of gingipain structural domains. All three gingipains possess a Sec signalling peptide, pro-peptide, catalytic domain and C-terminal domain. RgpA and Kgp also possess haemagglutinin/adhesin domains which are extensively processed by autolytic cleavage. Regions of sequence conservation are given the same colouring. Sites of proteolytic processing are indicated with an arrow and residue number. The HA domains are labelled according to the sites of autolytic processing in the *P. gingivalis* W50 strain [67]. (Adapted from [14, 70]).

RgpB is also found primarily associated with the cell surface with a predicted molecular mass of 56 kDa for the mature protein. Despite this Slakeski *et al.,* [77] observed migration of mature RgpB via SDS-PAGE as diffuse bands between 80-90 kDa. This membrane-associated form of RgpB was found to bind MAb 1B5, a monoclonal antibody raised against monomeric RgpA [78]. This antibody was also found to be immunoreactive against *P. gingivalis* lipopolysaccharide but not against soluble forms of the RgpA and RgpB or recombinantly expressed RgpA. It has since

been demonstrated that this antibody binds to the phosphorylated branched mannan in the APS repeating unit of A-LPS.

The crystal structure of soluble RgpB, solved in the same year, revealed that the mature protein has a C-terminal truncation at Ser^{435} (according to mature enzyme numbering; Ser^{664} according to full sequence numbering) [79]. It was therefore hypothesised that the gingipains are attached to the cell surface via C-terminal LPS attachment [67]. This led Veith *et al.,* [67] to perform proteomic analysis of outer membrane proteins present in *P. gingivalis*. They identified a group of proteins, including all three gingipains, HagA, P27 (PG1795) and P59 (PG2100/TapC) that were immunoreactive against MAb 1B5. It was also observed that all of these proteins possess a conserved C-terminal domain (CTD) with sequence similarity to the CTD of RgpB.

In 2006, Seers *et al.,* [80] discovered that removal of the CTD from RgpB caused accumulation of precursor isoforms in the periplasm and culture fluid, thus demonstrating the importance of this domain in gingipain export. The authors also performed sequence analysis of *P. gingivalis* open-reading frames to identify 34 putative CTD proteins, 24 of which had previously been observed to associate with the cell surface [80]. More recent analysis of 21 fully sequenced Bacteroidetes genomes has predicted the presence of 663 proteins possessing the conserved CTD [81]. The current list of known CTD protein encoded by *P. gingivalis,* some of which have yet to be experimentally verified, has been extended to 36 proteins; these are listed in Table 1.5.1.

Many of these CTD proteins have no known function, but some of them are associated with *P. gingivalis* virulence. For example, haem-binding protein 35 (Hbp35) is essential for *P. gingivalis* growth in haem depleted media and has been shown to bind thioredoxin and haemin [88]. *P. gingivalis* peptidylarginine deiminases (PPAD)

Table 1.5.1: Known CTD proteins in *P. gingivalis*.

W83 Locus Tag	ATCC 33277 Locus Tag	Protein Name	Description	References
PG0026	PGN_0022	PorU	T9SS cargo protein sortase	[82]
PG0182	PGN_0291	Mfa5	Minor fimbriae accessory protein	[83]
PG0183	-	-	Hypothetical CTD lipoprotein	[81, 84]
PG0232	PGN_0335	CPG70	Zinc carboxypeptidase	[67, 80, 82, 85]
PG0350	PGN_1611	-	Internalin-like protein	[80]*
PG0410	-	-	Hypothetical gingipain-like peptidase C25	[86]*
PG0411	PGN_1556	-	Putative haemmaglutinin	[80, 84]*
PG0495	PGN_1476	-	Hypothetical CTD protein	[80]*
PG0506	PGN_1466	RgpB	Arg-specific gingipain B	[67]
PG0553	PGN_1416	PepK	Lys-specific Serine protease	[80, 87]
PG0611	PGN_0654	-	Hypothetical CTD lipoprotein	[80, 82]*
PG0614	PGN_0657	-	Hypothetical CTD protein	[80]*
PG0616	PGN_0659	Hpb35	Haem binding protein	[80, 82, 88]
PG0626	-	Omp40	Hypothetical CTD protein	[67, 81]*
PG0654	PGN_0693	-	Hypothetical CTD protein	[80]*
PG0769	PGN_0795	-	Fibronectin type III domain protein	[81]†
PG0787	PGN_0810	-	Hypothetical CTD protein	[81]†
PG1030	PGN_1321	-	Hypothetical CTD protein	[80]*
PG1035	PGN_1317	-	Hypothetical CTD protein	[80, 89]
PG1326	PGN_1115	-	Putative haemmaglutinin	[80]*
PG1374	PGN_0852	PG97	Immunoreactive 47 kDa antigen	[82]*
PG1424	PGN_0898	PPAD	Peptidylarginine deiminase	[80, 82, 88]
PG1427	PGN_0900	Periodontain	Peptidase C10; PrtT-related	[76, 90]
PG1548	PGN_0561	prtT	Cysteine protease (domain peptidase C10)	[76, 91]
PG1604	PGN_0509	PorZ	β-propeller protein	[92]
PG1795	PGN_1770	P27	Hypothetical CTD protein	[67, 88]
PG1798	PGN_1767	PG99	Immunoreactive 46 kDa antigen	[80, 93]
PG1837	PGN_1733	HagA	Haemmaglutinin	[82]
PG1844	PGN_1728	Kgp	Lys-specific gingipain; Haemmaglutinin	[67, 80]
PG1969	-	-	Hypothetical CTD protein	[76, 81]*‡
PG2024	PGN_1970	RgpA/HagE	Arg-specific gingipain; Haemmaglutinin	[80]
PG2100	-	TapC/PG102	Hypothetical CTD protein	[80, 93]
PG2102	PGN_0152	TapA/PG91	Hypothetical CTD protein	[80, 93]
PG2172	PGN_0123	-	Hypothetical CTD protein	[82]*
PG2198	PGN_2065	-	Hypothetical CTD protein	[82]*
PG2216	PGN_2080	-	Hypothetical CTD protein	[80]

* Identified using proteomic methodology. MS used to identify CTD containing protein in the outer membrane.
† Identified using bioinformatic methodology only and not verified experimentally.
‡ No Sec signal peptide identified.

is a CTD protein capable of citrullinating (converting protein arginine to citrulline) host and endogenous proteins. This activity has been associated with protection from acid cleansing in the mouth [94], modulation of biofilm growth [95] and rheumatoid arthritis pathogenesis [96]. One of the more recently discovered CTD proteins is the minor fimbriae accessory component Mfa5 [83]. This protein is necessary for fimbriae polymerisation and mediates the incorporation of Mfa3 and Mfa4 into the minor fimbriae.

Alignment of the CTD sequences in *P. gingivalis* W83 revealed the presence of five conserved sequence motifs, labelled A-E [80]. Of these, motifs D and E, which comprise the 22 C-terminal residues, show the highest level of conservation. Furthermore, it was demonstrated that fusion of the 22 amino acid fragment is sufficient for the export of GFP from *P. gingivalis* [97]. This is consistent with motifs D and E having an important functional role in the export of surface-associated proteins from *P. gingivalis*. However, the overall sequence identity of the motifs is relatively low, which led to the suggestion that the domain is structurally, rather than sequentially, conserved [98].

1.6 Discovery of a System Related to Gingipain Export

Wild-type *P. gingivalis* grown on blood agar plates appear as black-pigmented colonies due to the accumulation of haem at the cell surface [99]. This pigmentation is dependent on Kgp activity and *P. gingivalis* mutants that are deficient in the surface expression of this protein grow as beige non-pigmenting colonies [100]. This relationship between pigmentation and Kgp activity makes it possible to use colony pigmentation as a screening tool for genes related to the export and maturation of

gingipains.

Initial studies used transposon mutagenesis for the production of non-pigmenting colonies. In these early studies, genes corresponded to the gingipains themselves, surface polysaccharide expression and the virulence modulating gene VimA, were identified [101–104]. Then, In 2005, this method was employed to detect the new gene, *porT*, that encodes a protein related to the export of gingipains [105]. In this mutant strain, gingipains were seen to accumulate in the periplasm rather than being exported to the cell surface. Furthermore, sequence analysis of the *porT* gene showed that it is only present in certain species of the *Bacteroidetes* phylum.

Following on from the discovery of *porT,* two more genes, *sov* and *PG0027*, were discovered which display very similar mutant phenotypes to that of the *porT* mutant [106, 107]. The name of the *PG0027* gene has since been changed to *lptO* [108] and is now more commonly referred to as *porV* [88]. In 2010, Sato *et al.,* [109] analysed the genomes of two *Bacteroidetes* species (*Cytophaga hutchinsonii* and *P. gingivalis*) that possess the *porT* gene compared to the genome of *Bacteroides thetaiotaomicron*, a *porT*-negative species of the *Bacteroidetes* phylum. This analysis yielded a list of 55 genes that were unique to *C. hutchinsonii* and *P. gingivalis*. Deletion mutants were created for 46 of these genes in *P. gingivalis*, eleven (including *sov* and *porT*) of which showed a decrease in the surface and extracellular gingipain activity. Five of the identified genes (*porK, porL, porM, porN, porP*) are located within a single operon in the *P. gingivalis* genome [110]. The other components identified in this study, and subsequently discovered genes, are located throughout the *P. gingivalis* genome. There are now 22 known genes that are associated with the export of gingipains and other CTD proteins from *P. gingivalis* (Table 1.6.1). Bioinformatic prediction has been used to predict an additional 11 proteins associated with CTD protein export but, no experimental data is available for these [111]. Therefore, they have not been

included in the table of system components.

This system was assumed to be novel because the identified proteins showed no significant sequence similarity to proteins of other known bacterial secretion systems. This novel system was named the Por secretion system (PorSS) [109] with the name later changed to the type IX secretion system (T9SS) to be consistent with standard nomenclature.

1.7 The T9SS Forms a Large Complex that Spans Both Membranes

When the PorK, L, M, N proteins were originally identified PorL and PorM were found primarily in the inner membrane as part of a 1.1 mDa complex, whereas PorK and PorN were found primarily localised to the outer membrane forming a 1.2 mDa complex [109]. In 2016, cryo-electron microscopy (cryo-EM) was used to visualise recombinantly expressed PorK and PorN [115]. A 50 nm diameter ring structure was observed comprising between 32-36 subunits of each protein which could form independent from any other system component. In recent analysis, using cryo-electron tomography (cryo-ET), it was confirmed that this structure also forms *in situ* (pre-print:[89]). The lipoprotein PorK is responsible for forming the upper portion of the ring and attaching the complex to the inner leaflet of the outer membrane, whereas PorN forms the lower portion of the ring complex. The ring complex was found to interact with the outer membrane β-barrel protein PorG and PorT [89, 115, 117]. The function of the ring complex, and associated PorG and PorT proteins is still unknown.

The interaction network of the PorKLMNP operon proteins was studied using bacterial two-hybrid and co-immunoprecipitation [117]. Their analysis revealed that PorL and PorM form a complex at the inner membrane which spans the periplasm to

Table 1.6.1: Known components of the T9SS in *P. gingivalis*.

W83 Locus Tag	ATCC 33277 Locus Tag	Protein Name	Description	Localisation	Interactions	References*
PG0026	PGN_0022	PorU	Cargo protein sortase	OM	PorV/Z	[82, 91]
PG0027	PGN_0023	PorV	β-Barrel; PorU anchoring and cargo protein shuttling	OM	Cargo Proteins; PorQ/U	[107, 112, 113]
PG0052	PGN_2001	PorY	Sensor histidine kinase	IM	porX; SigP	[109]
PG0162	PGN_0274	SigP	T9SS sigma factor	CP	PorX	[114]
PG0189	PGN_0297	PorG	β-Barrel	OM	PorK/N	[84, 115]
PG0192	PGN_0300	Omp17	β-Barrel	OM	-	[116]
PG0287	PGN_1677	PorP	β-Barrel	OM	PorE/K/M; PG1035	[109, 117]
PG0288	PGN_1676	PorK	Lipoprotein; forms ring complex with PorN	PP	PorG/M/N/P/T/W; Sov	[89, 109, 115, 117]
PG0289	PGN_1675	PorL	Trimeric inner membrane protein	IM	PorM/X	[109, 117]
PG0290	PGN_1674	PorM	Dimeric inner membrane protein	IM	PorK/L/N/P	[109, 117, 118]
PG0291	PGN_1673	PorN	Forms ring complex with PorK	PP	PorG/K/M/T/W; Sov	[109, 115, 117]
PG0534	PGN_1437	PorF	β-Barrel; TonB dependent receptor	OM	-	[119]
PG0602	PGN_0645	PorQ	β-Barrel; Anchors PorZ to surface	OM	PorV/Z	[109, 112]
PG0751	PGN_0778	PorT	β-Barrel	OM	PorK/N	[105]
PG0809	PGN_0832	Sov	β-Barrel; Secretin	OM	PorK/N/V/W	[106, 113]
PG0928	PGN_1019	PorX	Chemotaxis protein	CP	PorL/Y; SigP	[109, 120]
PG1058	PGN_1296	PorE	Lipoprotein; Cell wall anchoring	PP	PorP; Peptidoglycan	[111, 121]
PG1604	PGN_0509	PorZ	β-Propeller; A-LPS attachment	OM	PorQ/U	[76, 92]
PG1850	PGN_1783	-	periplasmic protein	PP	PorW	[89, 111]
PG1947	PGN_1877	PorW	Lipoprotein	PP	Sov; PG1850	[109]
PG2092	PGN_0144	Plug	Sov channel plug	PP	Sov	[113]

*Original citations for protein identification and key citations with novel data pertaining to protein function, structure and/or interactions are listed.

interact with the KN ring via the periplasmic portion of PorM. The outer membrane β-barrel protein PorP was also shown to form part of this complex by binding to PorK and PorM. In 2018, the periplasmic portion of PorM was crystallised and the structure resolved using X-ray crystallography [118]. The periplasmic portion of PorM is formed of four similar domains in an extended arrangement with a ~45° angle between domains D2 and D3 and a total length of ~180 Å, capable of spanning most of the periplasmic space [118]. These results are consistent with the observation that PorM bridges the inner and outer membrane complex of the T9SS. It has recently been demonstrated, in *Flavobacterium johnsoniae*, that the transmembrane helices of each GldM (PorM) monomer interact with five GldL monomers (pre-print: [122]). Each GldL monomer possesses two transmembrane helices that form a ring structure around the transmembrane helices of GldM, with the N- and C- termini of GldL localising to the cytoplasm. Given that GldM and GldL possess conserved glutamate residues within the hydrophobic transmembrane helices it was speculated that these two proteins may be responsible for energy transduction in the T9SS [123]. To test this, single residue mutants (E49Q and E49D) were created for the conserved glutamate residues in GldL. These mutants showed a significant reduction in gliding motility and chitinase activity, two phenotypes that are dependent on a functional T9SS in *F. johnsoniae* (pre-print:[89]).

Another recent study has used co-immunoprecipitation to understand the interactions that exist between components of the T9SS (pre-print:[89]). The periplasmic lipoprotein PorE was found to form a complex with PorP and the CTD protein PG1035, predicted to bind to the outer face of PorP. PorE was originally discovered using comparative genome analysis and was predicted to comprise four domains: a tetratricopeptide repeat (TPR) domain, a β-propeller domain, a carboxypeptidase regulatory domain-like fold (CRD) and an OmpA_C-like putative peptidoglycan binding

domain [111]. A crystal structure of the C-terminal OmpA_C-like domain has been solved [121]. This structure possesses a protein cleft similar to that of other peptidoglycan binding proteins and it was possible to model a peptidoglycan fragment within the cleft. This data suggests that PorE may be responsible for anchoring the T9SS to the bacterial cell wall. However, the function of the other three domains is unknown. PorF is predicted to be a β-barrel protein that has been shown to localises to the outer membrane in *P. gingivalis* [119]. However, no function has been assigned to this protein, nor has it been identified in any of the studies analysis protein-protein interactions of T9SS components.

1.8 Sov Mediates Export of Cargo Proteins Across the Outer Membrane

In 2007 the outer membrane protein Sov was identified and found to be essential for T9SS functionality [106]. Given the length of the translated peptide sequence (~2500 aa) it was speculated that Sov may form an outer membrane channel of the size necessary to transport folded proteins across the membrane. In 2018, Lauber *et al.,* [113] used cryoEM to observe the *F. johnsoniae* homolog of Sov, SprA.

The resulting structure was of a 36 strand β-barrel protein that is capped at the extracellular face but with a lateral opening to the surface of the outer membrane. Two other proteins were found in complex with SprA. The first is a novel lipoprotein peptidyl-prolyl cis-trans isomerase (PPI) that did not appear to have any effect on T9SS function. The second protein identified was PorV, bound to the lateral opening of SprA, with three extracellular loops dipping into the central pore. Purification of SprA following the removal of the *porV* gene led to SprA being purified with another novel protein, simply called the Plug protein. This name was given because it was

found to occlude the central pore of SprA by binding at the periplasmic face. The model of these two complexes suggests that there is a gating between a plug bound and PorV bound state to allow regulation of the system and prevent aberrant export of periplasmic proteins and components.

Blue native (BN)-PAGE of Sov in *P. gingivalis* has revealed that it also interacts with the periplasmic lipoprotein PorW, which in turn interacts with PGN_1783 (PG1850) (preprint: [89]). PorW is predicted to possess several TPR domains, which are commonly involved in protein-protein interactions. It may therefore have a role in attaching to other T9SS components or cargo proteins present in the periplasm. The authors also demonstrate that Sov complexes with PorN, suggesting that the Sov complex directly interacts with the KN ring.

1.9 The T9SS is Linked to LPS Transport

During the initial studies of the *porV* gene, it was found to have a similar phenotype to the *porT* deletion mutant [107]. Further analysis of these two mutants revealed that A-LPS accumulates within the periplasm along with unprocessed CTD proteins [108]. The lipid profile of the PorT mutant was found to be similar to that of the wild-type. However, the PorV mutant was unable to deacylate penta-acylated Lipid A to tetra-acylated Lipid A. In the PorV mutant high levels of non-phosphorylated Lipid A and novel forms of Lipid A, not seen in the wild type, were also observed. It has also been demonstrated that disruption of A-LPS synthesis causes CTD proteins to accumulate in the periplasm [124]. It was suggested that the secretion of A-LPS and CTD proteins is co-ordinated and that disruption of either process causes blockage of the entire system.

The *porU* gene, identified by comparative genome analysis, is located directly upstream of *porV* and was previously found to be essential for gingipains secretion

[109]. Bioinformatic analysis of the protein sequence revealed that it has a putative CTD and similarity to the catalytic domain of the gingipains with His^{657} and Cys^{690}as catalytic residues [82]. Due to the presence of a CTD in PorU, the protein was assessed for outer membrane localisation and LPS modification. Despite localising to the outer membrane the PorU protein was not modified via A-LPS attachment, nor was the CTD cleaved from the mature protein.

Analysis of the secretion profiles in a *porU* deletion mutant and catalytic mutant (C690A) showed that there is only partial secretion of gingipains in these strains and some accumulation of CTD proteins in the periplasm [82]. Despite detecting CTD proteins at the cell surface, no LPS modification or CTD removal of cargo proteins occurred. Analysis of RgpB maturation showed that processing was stalled at the stage of CTD removal in the two mutant strains, suggesting that PorU is responsible for the proteolytic cleavage of the cargo protein CTD.

A *wbaP* mutant strain of *P. gingivalis*, which does not produce A-LPS, showed an accumulation of T9SS cargo proteins in the culture fluid with cleaved CTDs but no LPS modification [91]. Mass spectrometry analysis of these released proteins revealed that they are modified at the C-terminus via the attachment of single amino acids or short peptides, suggesting that C-terminus is the site of cargo protein modification. This was confirmed by analysis in the wild-type, where CTD proteins were found to be attached to a 648 Da linker molecule via a peptide bond. Based on this evidence it was suggested that PorU is responsible for the conjugated removal of the CTD and C-terminal modification of T9SS cargo proteins via a sortase-like mechanism [91].

In 2014, purification of tagged PorV from *P. gingivalis* showed that it co-purifies with PorU, suggesting that is responsible for anchoring PorU to the cell surface [125]. Further analysis of this complex by BN-PAGE revealed the presence of a 440 kDa complex comprising PorU, PorV, PorZ and PorQ and smaller sub-complexes of PorU:Porv

and PorZ:PorQ [112]. PorZ is similar to PorU, in that it possesses a CTD that is not removed following transfer to the outer membrane [92]. It is likely that the CTD of PorU and PorZ mediates anchoring to PorV and PorQ, respectively. Deletion of PorU causes the accumulation of immature cargo proteins bound to PorV and A-LPS bound to PorZ [112]. From this data it was suggested that PorV acts as a shuttle protein delivering CTD proteins to the 'attachment complex', composed of PorU, PorV, PorZ and PorQ, for CTD removal and LPS attachment.

The outer membrane Omp17 was found to be necessary for CTD cleavage and surface attachment of cargo proteins [116]. Deletion of the *omp17* gene caused CTD proteins to be released into the culture fluid and is necessary for surface expression of PorU. However, the exact role of Omp17 in this process is unclear.

1.10 Regulation of the T9SS

PorX and PorY encode a two-component regulatory system involved in the regulation of the T9SS [109]. The PorX response regulator is the only known component of the T9SS to fully localise to the cytoplasm. The sensor kinase PorY is localised to the inner membrane via two transmembrane helices [120]. Recombinantly expressed PorY is able to auto-phosphorylate and transfer a phosphoryl group to PorX. PorX is then able to activate the extra-cytoplasmic function sigma factor SigP which can bind to the promoter region of T9SS genes [114]. PorX has also been shown to interact with the cytoplasmic domain of PorL and may transfer regulatory information to the PorL/M motor complex. However, the specific environmental factors that affect this system are not known.

1.11 Current Model of T9SS Secretion

Evidence for the structural arrangement and functional role of T9SS components is increasing at a fast rate. When work for this project was begun there was no structural data available for any system component and very limited understanding of individual protein function. However, it is now possible to create a much clearer picture of how cargo CTD proteins are exported and processed by the T9SS.

There are essentially two major steps during the secretion process, summarised in Figure 1.11.1. The first step involves recognition of cargo proteins, that have passed into the periplasm via the Sec system, and exporting them to the outer membrane via Sov. There is also evidence that the CTD proteins interact with multiple periplasmic domains of PorM and are then passed to PorN (preprint: [126]). It is not clear if there are interactions between CTD proteins and system components prior to recognition by PorM, or in between PorN binding and transfer to the Sov complex. The second stage of the secretion process involves shuttling of cargo proteins, via PorV, to the attachment complex where the CTD is removed and the A-LPS modification occurs. At this stage the cleaved CTD is released from the cell.

Understanding of how these complexes interact is also developing. The KN ring appears to facilitate the assembly of multiple protein complexes involved in secretion including the PorL/M motor complex and the Sov complex. However, it is not clear whether the complexes that associate with the KN ring are localised to the inside or outside of the ring, or if they are free to pass in and out. There is no data to suggest the attachment complex directly interacts with the KN ring and evidence from BN-PAGE analysis suggests that the attachment complex is independent from other proteins of the T9SS. PorF and Omp17 remain the only proteins that have experimental data supporting their role in the secretion process that are not known to interact with any of these complexes.

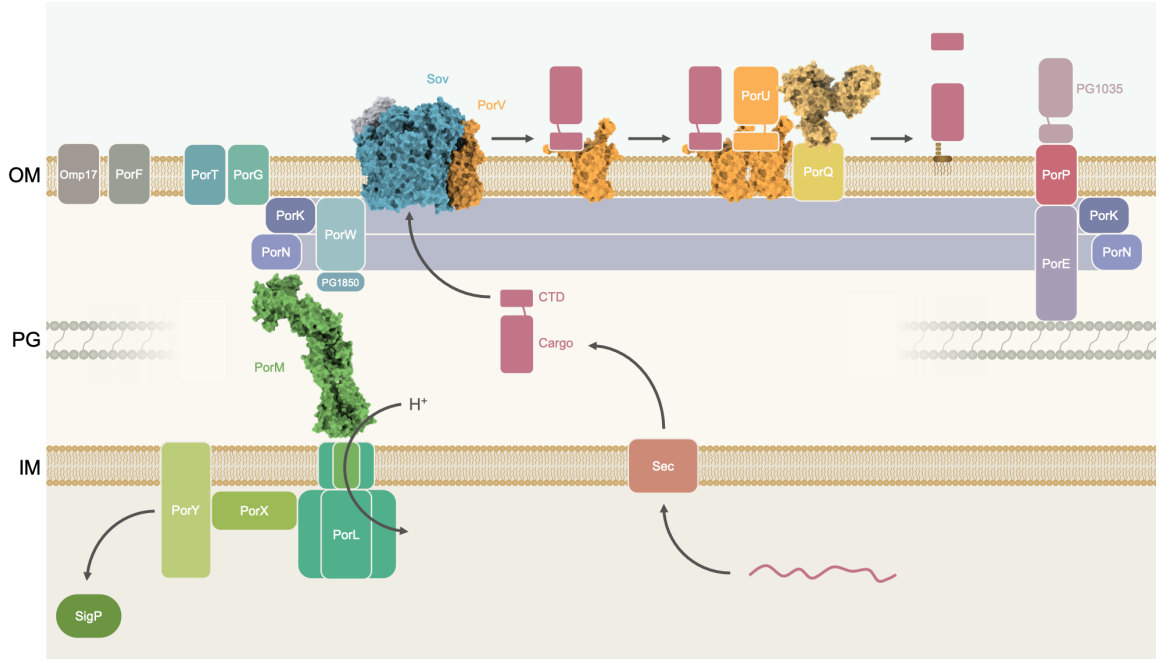


Figure 1.11.1: Schematic model of the *P. gingivalis* T9SS. T9SS cargo proteins possess a N-terminal signal peptide, for transfer through the Sec system into the periplasm, and a CTD, for transfer through the T9SS. Periplasmic cargo proteins are recognised by the T9SS and pass to Sov, the outer membrane translocon. The cargo protein is then transferred to PorV which then shuttles the cargo protein to the attachment complex, comprised of PorQ, U, V and Z. Finally, the PorU sortase cleaves o the cargo CTD and adds an A-LPS molecule to the C-terminus of the cargo anchoring it to the membrane. The PorL/M complex is believed to power export of cargo protein across the outer membrane using proton motive force. The two-component system, comprised of PorX and PorY, is involved in regulation of the system via interaction with the sigma factor SigP and porL. The PorK/N ring complex forms an anchor for many other proteins of the T9SS to bind. PorE interacts with the PorP:PG1035 complex in the OM and anchors the system to the peptidoglycan layer. PorT and Omp17 are not known to interact with any other component of the T9SS. Structural models are shown for proteins that have an experimentally determined atomic structural model available. Legend: IM - Inner membrane, PG - Peptidoglycan, OM - Outer membrane.

1.12 Project Aim

1.12.1 Generation of a Structural Model for the T9SS CTD

The first aim of this project was to develop a structural model for the CTD of T9SS cargo proteins. When work on this project was started there was no available structural model for the CTD of any T9SS cargo. Given that there is low levels of sequence identity between the CTD of cargo proteins it is possible that the information conferring recognition by the T9SS is contained within the secondary and tertiary structure rather than the primary amino acid sequence. It was therefore reasoned that solving a structural model of the CTD would significantly advance our understanding of how cargo proteins are recognised by the T9SS.

1.12.2 Identification of T9SS Components that Interact with Cargo Protein CTD

When this project was conceived there was no available information regarding the interaction of T9SS proteins and CTD proteins. Indeed, most of the identified components still had no functional characterisation beyond the effect of gene deletion of the secretion profile and surface protease activity in *P. gingivalis* [127]. Therefore, the second of aim of this project was to identify components of the T9SS that are capable of interacting with the CTD of cargo proteins. This aim was selected as it would help to elucidate the functional role of T9SS proteins and provide a pool of interaction partners which could be investigated further.

1.12.3 Characterisation of T9SS Interaction with the Cargo Protein CTD

The final aim of this project was to characterise the most promising identified interaction(s) in further detail, with the ultimate aim of generating a structural model for this complex. To date, no structural models are available for a CTD protein in complex with any T9SS component. The characterisation of this interaction and generation a structural complex would further our understanding of how CTD proteins interact with the T9SS and how the system mediates protein transport. This may have also led to the identification of important structural motifs in the CTD that explain how this domain is recognised by the T9SS.

Chapter 2

Theory of NMR Spectroscopy

The method of nuclear magnetic resonance (NMR) was first developed in 1946 through the experiments of Bloch and Purcell [128, 129] and is now widely used by physicists, chemists and biologists. In this project, NMR has been the primary method used for studying protein structure and dynamics. Furthermore, unlike other techniques described in this thesis, NMR spectroscopy is the least intuitive technique to understand and interpret. Therefore, this chapter will cover the basics of NMR theory and how it can be applied to the study of biomolecules such as proteins. Much of the theoretical descriptions that follow are derived from the book 'Understanding NMR spectroscopy' by James Keeler [130], unless referenced otherwise.

2.1 Principles of Nuclear Magnetism

2.1.1 Nuclear Spin and Spin States

The subatomic nuclei of atoms possess a property known as nuclear spin, an intrinsic type of angular momentum. This spin property is characterised by the spin quantum number, *I,* which can take either whole or half-integer values. As a general rule, most

atomic particles that have an odd mass number will have half-integer I values and are called spin-half nuclei. For particles where $I = 0$, the spins of each nucleon interact to cancel out. Therefore, such nuclei are termed to be NMR invisible. However, it is possible to use nuclear isotopes such that the particles necessary for the characterisations of proteins, primarily ^{13}C and ^{15}N, take a half-integer I value.

Spin-half nuclei, act as small bar magnets, with a nuclear magnetic moment denoted by the symbol, μ. The magnetic moment, defined in Equation 2.1.1, is proportional to the spin of the nuclei, *I,* and an intrinsic property of the nuclei called the gyromagnetic ratio (γ). γ is constant for a specific type of nucleus and is defined as the ratio of the magnetic moment to the angular momentum of the particle. The gyromagnetic ratios for nuclei commonly detected in protein NMR experiments are shown in Table 2.1.1.

$$\mu = \gamma I \tag{2.1.1}$$

Table 2.1.1: Spin properties of isotopes commonly used in biomolecular NMR.

Isotope		**Spin**	**Gyromagnetic Ratio (10^7 T^{-1} s^{-1})**	**NMR Frequency at 14.093 Tesla (MHz)**	**Chemical Shift Range (ppm)**
1	H	1/2	26.752	600.000	-1 to 12
2	H	1	4.107	92.104	-1 to 12
13	C	1/2	6.728	150.870	0 to 200
15	N	1/2	-27.13	60.821	0 to 900

2.1.2 Bulk Magnetisation

When analysing samples by NMR spectroscopy we consider not the individual magnetic moments, but the net or bulk, magnetisation of the sample. Bulk magnetisation is denoted by the symbol M and defined as the sum of each magnetic moment. In the absence of a magnetic field, spin orientation is random and the magnetic moments essentially cancel each other out (Figure 2.1.1A). In the presence of a static magnetic field (B_0) there are two possible energy levels the spins can adopt; a high energy and a low energy state. These energy levels are described by the magnetic quantum number (m), which for the case of spin-half nuclei can be either $m = -\frac{1}{2}$ or $m = +\frac{1}{2}$. The low-energy state ($m = -\frac{1}{2}$) is termed as the α-state and the high energy state ($m = +\frac{1}{2}$) is termed the β-state.

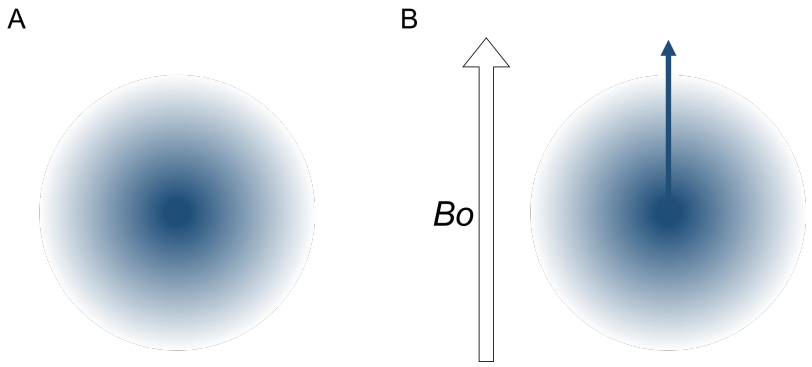


Figure 2.1.1: The distribution of spin orientation in an external magnetic. (A) In the absence of an external magnetic field, the orientation of the magnetic moments is random, represented by the cloud of possible distributions. (B) When the sample is placed in an external magnetic field (B_0) the spins are still orientated in all directions but now have a slight propensity to align with the external field. This magnetisation vector (M) is shown by the large vertical arrow aligned along the z-axis of the B_0 field.

It is often misrepresented that the α and β states relate to parallel and antiparallel alignment of the spins relative to B_0, respectively. In reality, spins still orientate in an almost entirely random manner in the presence of the B_0 field due to the thermal energy of the sample [131]. There is, however, a slight propensity for the spins to point

along the direction of the B_0 field (defined as the z-axis). This alignment, although slight, gives rise to bulk magnetisation of the sample along the z-axis, represented by the blue arrow in Figure 2.1.1B. This magnetisation vector is central to the technique of NMR spectroscopy and all following discussions will be in regard to this vector.

2.1.3 Larmor Precession

In a static magnetic field, in which the magnetisation vector is aligned with B_0, the vector product of M and B_0 will be zero. Therefore, the magnetisation vector will be static. However, if there is a non-parallel alignment of M relative to B_0, then M will precess around B_0 at the angular frequency, commonly referred to as the Larmor frequency, of ω_0 (rad s^{-1}) given in Equation 2.1.2.

$$\omega_0 = -\gamma B_0 \tag{2.1.2}$$

This equation shows the frequency of precession is dependent on field strength and the gyromagnetic ratio (γ) of the nucleus. The Larmor frequency for commonly detected nuclei in a 14.093 Tesla field are shown in Table 2.1.1. We can see from this that the frequency of a ^{1}H spin is much higher than that of ^{13}C and ^{15}N spins. Due to the higher frequency at which ^{1}H spins precess they give a sharper signal with a better signal-to-noise ratio. As such, NMR spectra are often detected using proton magnetisation.

2.1.4 Chemical Shift and Shielding

In NMR the Larmor frequency is dependent on field strength (Equation 2.1.2). This makes it confusing to express characteristic resonances in Hz because nuclei will precess at different frequencies depending on spectrometric power. Therefore, precession frequency is written in ppm relative to a chemical standard, as derived in Equation

2.1.3. Where, ν is the resonance frequency for the nuclei of interest and ν_0 is the resonance frequency of the reference compound. When represented in this way, the frequency is referred to as the chemical shift of the magnetic moment.

$$\delta(ppm) = 10^6 \cdot \frac{\nu - \nu_0}{\nu_0} \tag{2.1.3}$$

Chemical shift values vary for nuclei of the same type due to an effect called shielding. Shielding is caused by the electrons surrounding the nucleus, which are induced by the external magnetic field (B_0) to produce local induced fields (B_{Local}). Therefore, the actual magnetic field experienced by the nucleus is the sum of these two fields [132]. This shielding effect causes differences to be observed in chemical shift that is dependent on the chemical and the structural environment of the nuclei.

Several chemical shift references are available and commonly used. In protein NMR spectroscopy the residual H_2O signal is commonly used for referencing to avoid any possible interference of the reference with the sample [133]. This was the referencing method used in this study. However, the reference chemical 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was used as an external standard, to calibrate the spectrometer for the temperature-dependent water signal.

2.1.5 Relaxation

In NMR spectroscopy the Larmor precession is induced by the use of radio frequency (RF) pulses perpendicular to *Bo*. However, when this applied field is removed, transverse magnetisation is lost as the spins return to equilibrium, a process referred to as relaxation. It is caused by the interaction of individual magnetic moments with local magnetic fields in the sample (B_{Local}). The type of local field that has the most impact arises from other non-zero spin nuclei, in particular ^{1}H nuclei. As such, the rate of relaxation varies between nuclei, due to their proximity to other nuclei in the system.

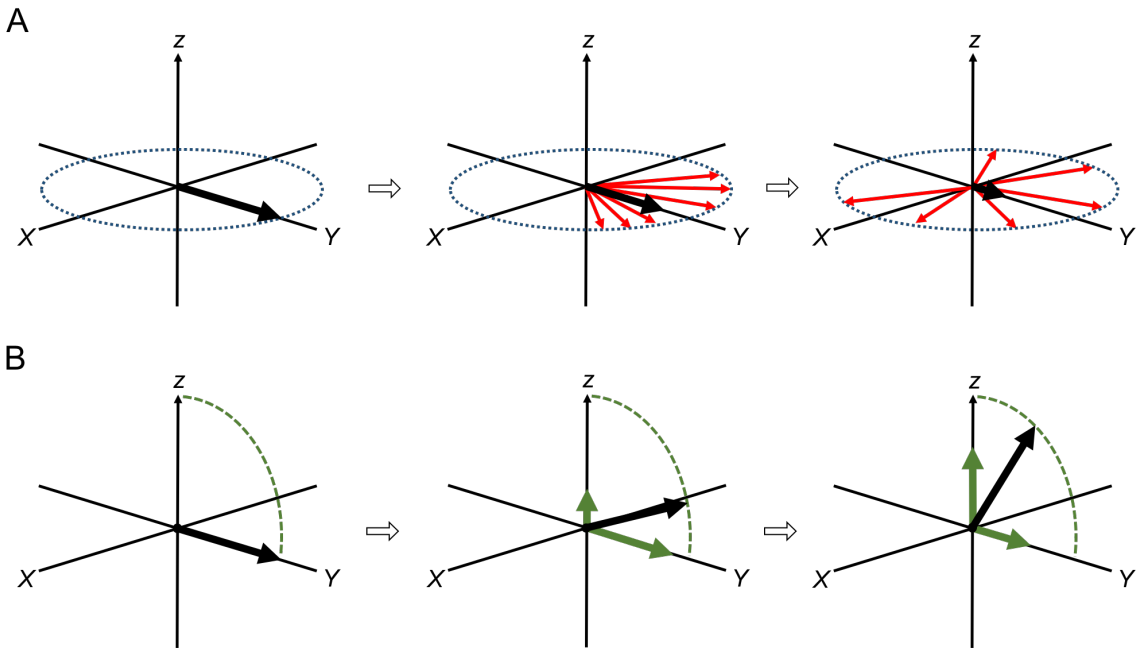


Figure 2.1.2: Transverse and longitudinal relaxation of magnetisation. (A) Transverse relaxation occurs due to interactions with local magnetic fields which alter the frequency of each spin precession. The red arrows represent individual magnetic moments that de-phase over time. This leads to loss of coherence and thus the magnetisation vector is diminished, represented by the large black arrow. (B) Longitudinal relaxation occurs due to re-equilibration of the individual magnetic moments. The green arrows represent the longitudinal and transverse components of magnetisation as M moves back to the z-axis.

There are two types of relaxation, both of which are driven by local magnetic fields as described above. The first is that of transverse relaxation, also referred to as spin-spin relaxation, in which there is a loss of sample coherence in the transverse plane. The B_{Local} field interacts with the magnetic moments to change the frequency of precession, as shown in Equation 2.1.4. As each nuclei experiences a different B_{Local} field, the Larmor precession will de-phase over time and spin coherence will be lost. When this occurs the magnetic moments cancel each other out and net transverse magnetisation is diminished, as shown in Figure 2.1.2A. The time constant

for transverse relaxation is referred to as transverse relaxation time (T_2).

$$\omega_0 = -\gamma\,(Bo + B_{Local}) \tag{2.1.4}$$

The second type of relaxation is known as longitudinal, or spin-lattice, relaxation and involves the re-equilibration of the magnetisation vector. Re-equilibration refers to the movement of M from the transverse plane back to being aligned with the B_0 static field. This process causes a decrease in transverse magnetisation and a relative increase in longitudinal magnetisation until M is realigned with B_0, as shown in Figure 2.1.2B. The time constant for this is referred to as longitudinal relaxation time (T_1).

2.1.6 Spin Coupling

2.1.6.1 Scalar coupling

Scalar coupling, or J coupling as it is often referred to, is the indirect interaction between nuclear magnetic moments mediates via the shared electron cloud. The interaction between σ bonding orbital electron spins creates an effective nuclear coupling that can transverse through the length of several bonds [134, 135]. For two coupled spins, A and B, the effect experienced by spin A will change depending on the quantum state of spin B. This difference causes the peak to split into a multiplet, as shown in Figure 2.1.3.

This effect is independent of magnetic field strength and is therefore represented as a frequency in Hz. The symbol that denotes J-coupling is ${}^{\mathrm{N}}\mathrm{J}_{\mathrm{A,\,B}}$ where A and B are the spins involved in the coupling and N is the number of bonds between the two spins. Geminal coupling refers to an interaction between two bonds and vicinal coupling refers to an interaction through three bonds. Coupling over greater distances is possible but is uncommon.

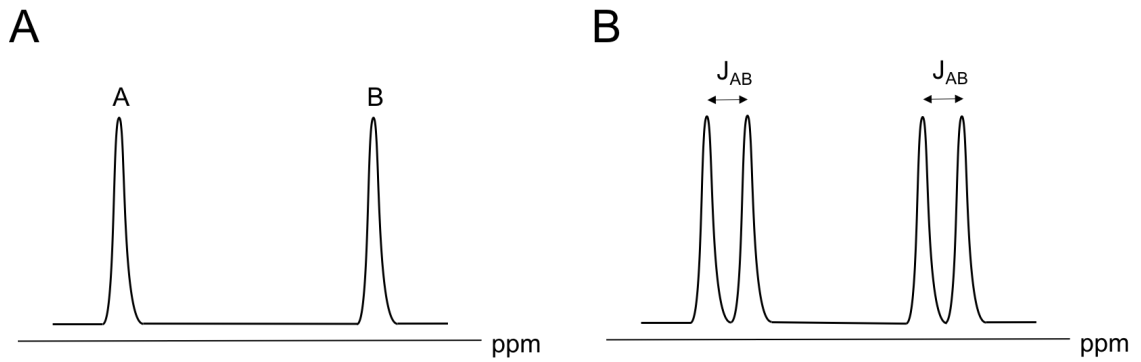


Figure 2.1.3: J-Coupling in a two spin system. (A) Two uncoupled spins, A and B, will give two single peaks in a 1D NMR spectrum. (B) If spins A and B are J coupled they will give doublet peaks in an NMR spectrum. The doublet peaks are separated by the coupling constant J_{AB}.

J coupling is central to bio-molecular NMR as it is through J-coupling that magnetisation is transferred in many of the more advanced experiments. Furthermore, it can be used to derive structural information about the sample. For example, vicinal J coupling in proteins can provide information on ϕ and φ backbone dihedral angles [136] and side-chain χ_1 angles [132].

2.1.6.2 Dipolar Coupling and the Nuclear Overhauser Effect

The second type of spin interaction between two magnetic moments is dipolar coupling. Unlike J-coupling it occurs through space, rather than covalent bonding. Dipolar coupling cannot be directly detected by NMR spectroscopy due to the isotropic tumbling of molecules in solution [137]. This tumbling causes the dipolar couplings to average to zero, cancelling the effect. However, we can detect dipolar couplings indirectly via cross-relaxation, also known as the nuclear Overhauser effect (NOE). The NOE was first described in 1953 [138] and refers to the change in peak intensity of a given spin when its dipolar coupled partner is irradiated to saturation (see Section 2.2.4). Saturation of the source spin causes cross-relaxation of the coupled spin and leads to a change in signal intensity. The magnitude of this change is directly related

to the inter-spin distance. The NOE is restricted to short-range interactions of ~5 Å because of the distance dependence of $1/r^6$ for the dipole-dipole coupling where, r is the intermolecular distance [132]. Since the strength of the effect is dependent on the distance between the interacting spins it can be utilised to determine a structural map of the atoms within a molecule.

2.2 Making use of the Larmor Precession

2.2.1 Radio Frequency Pulses

Larmor precession of the magnetisation vector is the observable signal in NMR spectroscopy. As the vector precesses an oscillating current is generated in a receiver coil that is aligned in the transverse-plane relative to the sample. The receiver coil detects only the transverse-component of magnetisation and none of the longitudinal-component. Relaxation causes the signal to decay over time, giving rise to the name free induction decay (FID).

The Larmor precession can be induced by the use of radio frequency (RF) pulses which push the magnetisation vector away from the z-axis. However, it is not a trivial matter to overcome the strength of the Bo field; we cannot simply apply a stronger field in the direction of the x-axis. Instead, a much weaker RF field is applied that exploits the phenomenon of resonance to overcome the strength of the B_0 field. An RF field that is on resonance with the Larmor frequency can be used to tilt the magnetisation vector away from Bo.

The magnetisation vector now experiences the vector sum of the static field, B_0, and the applied field, B_1. The duration of the RF pulse, for time t_p, governs the angle to which the magnetisation vector is rotated. This is called the flip angle, denoted by the symbol β and given in Equation 2.2.1, where ω_1 is the precession frequency

around the B_1 field. It is common to use pulses giving flip angles of $\pi/2$, which is a 90° excitation pulse, and π, which is a 180° inversion pulse.

$$\beta = \omega_1 t_p \tag{2.2.1}$$

When we have a system with multiple spins, each with a different precession frequency, our RF pulse will not resonate with all of them. However, if the RF pulse is strong enough we can flip spins that have precession frequencies several kHz away from the frequency of the RF pulse. This so-called hard pulse is achieved by selecting a large bandwidth and it is used to flip all spins of the system.

2.2.2 1-Dimensional NMR

By use of a hard pulse, we can now record the most basic 1-dimensional experiment, consisting of three steps. 1, the spins are allowed to come to equilibrium, 2, the RF transmitter flips the spins by 90° with a hard excitation pulse, and 3, the RF transmitter is switched off and the resultant FID is recorded. These stages are represented in the commonly used pulse pattern diagram (Figure 2.2.1), in which the excitation pulse is shown as a black rectangle and the acquisition is represented by a decaying sine wave. This process can then be repeated many times to remove residual background signals and improve the signal to noise.

Although 1D NMR has limited use in the resolving of protein structures it can provide useful information as to the stability, solubility, degree of folding and purity of a protein sample. Such information provides a useful guide before pursuing the expensive preparation techniques involved in expression and purification of isotopically labelled protein samples.

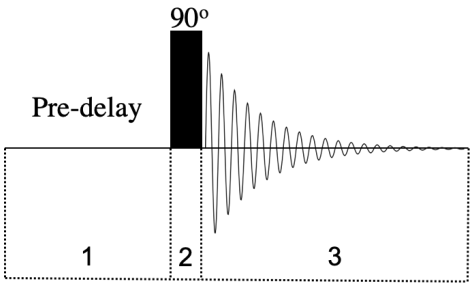


Figure 2.2.1: 1D pulse pattern. In step 1 the pre-delay allows spins to come to equilibrium and recover from any previous experiments. In step 2 a hard pulse is used to tilt the magnetisation by 90° . In step 3 the free induction decay signal is detected and recorded.

2.2.3 Spin-Echo

The spin-echo was first developed in 1950 by Erwin Hahn [139] and is now a key building block of multi-dimensional experiments. The spin-echo pulse sequence has the interesting property of re-phasing, or 'refocusing', the spin magnetisation as shown in the phase diagram of Figure 2.2.2. After magnetisation is transferred to the transverse plane with an excitation pulse, the different precession frequencies of each spin causes dephasing. After a set time τ, a 180° pulse is applied which flips the magnetisation in the transverse plane, i.e. it is moved from *xy* to -*x*-*y*. After a second time of τ, the spins will be rephased because of the 180° inversion pulse. This can be seen in the phase diagram of Figure 2.2.2 for two spins with different precession frequencies. Crucially, this effect works regardless of precession frequency and for different times of τ.

One reason we use this sequence in NMR is to remove any artefacts or 'acoustic ringing' from our initial pulse, particularly from low-frequency nuclei such as ^{15}N. The delay caused by the spin-echo prior to data acquisition provides time for this unwanted signal to decay. The second use of the spin-echo is to rephase signals that we want to detect whilst ignoring those that we don't. The most simple example of this is water suppression, in which the strong water signal is reduced to give clearer spectra with

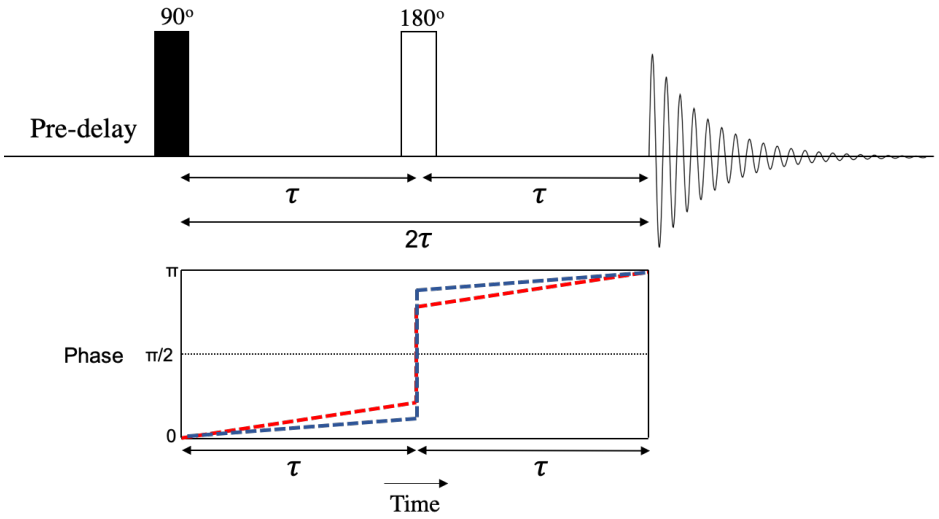


Figure 2.2.2: Spin-Echo Pulse Pattern. The precession of two spins is shown with the red and blue lines. Using the 180° inversion pulse the phase is re-aligned for the two spins after the 2τ time period.

a better signal to noise. To achieve this the spin-echo inversion pulse is selectively applied to the whole sample excluding water signal frequency. This leaves the water signal dephased and it cancels out during data acquisition. This technique can be applied to selectively include and exclude any frequency ranges in the sample.

2.2.4 Saturation and Decoupling

Although J-coupling provides useful structural information, the multiplet peaks can complicate data analysis, particularly for complex systems. Additionally, the split multiplet peaks have less intensity giving a lower signal-to-noise ratio. To overcome these issues the spins can be 'decoupled' to give a single intense peak in the spectrum.

Decoupling can be achieved by irradiating the sample such that the population of spins in the α and β energy states are equal. This process is referred to as saturation. Saturation causes loss of the NMR signal because the two equal populations cancel each other out ($Mx = My = Mz = 0$). Broadband decoupling uses a series of weak pulses that target a wide range of spins; usually targets specific nuclei type such

as ^{15}N or ^{13}C. Increasing the RF pulse length at a lower power enables a narrower range of spins to be targeted, such as α- or carbonyl-carbons. These broadband pulse sequences are often applied during the detection period to heteronuclei (e.g. ^{15}N or ^{13}C) to decouple them from the nuclei being detected (e.g. ^{1}H).

A variation of the spin-echo can also be applied to decouple spins during the course of the experiment. In a heteronuclear experiment, involving ^{1}H and ^{15}N spins, applying a spin-echo to both spins will refocus the phase of both spins. However, if the 180° pulse is targeted only to the ^{1}H nuclei, then only these nuclei will be refocused and not the ^{15}N nuclei. This causes the effect of the coupling to be refocused which effectively decouples the ^{15}N spins from the ^{1}H spins.

By combining the spin-echo decoupling and broadband decoupling techniques, spins in a heteronuclear system can be decoupled to reduce the peak multiplet into a single peak with greater intensity.

2.2.5 Two-Dimensional NMR

2-Dimensional NMR was first developed during the late-1970's and is now routinely used in the fields of chemistry and biochemistry [140, 141]. It is particularly useful for studying complex systems as it increases the resolution of each peak. That is, a collection of peaks that may be severely overlapped in a 1-dimension can be separated by addition of a second dimension.

In 2D NMR there are four basic stages to the experiment: preparation, evolution, mixing and detection. These steps are illustrated in Figure (2.2.3A). During the preparation period, spins are first brought to equilibrium and then excited by one or more pulses. With the spins now precessing, magnetisation is allowed to evolve for the variable time period, t_1. During the mixing period, magnetisation is then transferred from one spin to another via scalar or dipolar coupling. In some cases, the type of

magnetisation present during the evolution period is not a detectable signal. The mixing period can therefore be used to convert magnetisation back into a detectable signal. Finally, the detection period, labelled t_2, is when the FID is acquired.

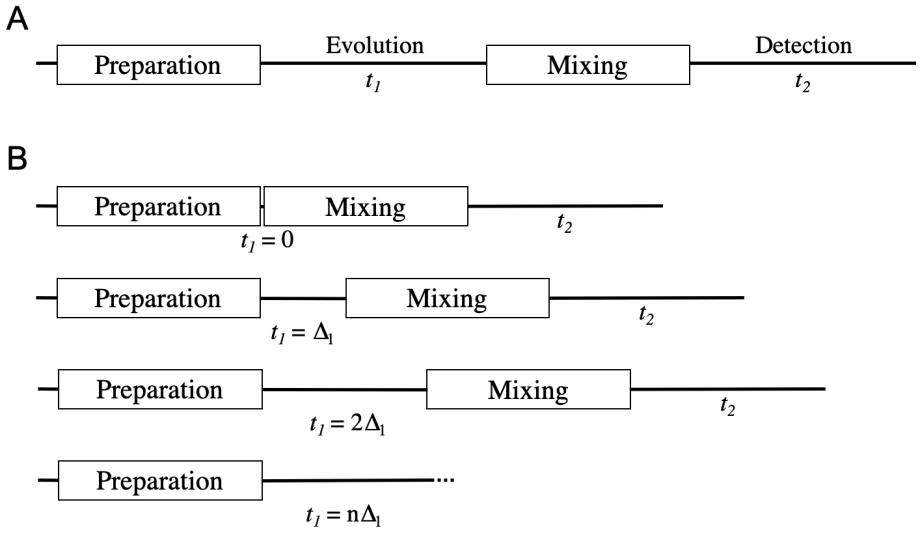


Figure 2.2.3: Schematic of stages in 2D pulse pattern. (A) Schematic of the four basic stages in 2D NMR. (B) Schematic of how t_1 evolution time is varied to measure different frequencies in the F_1 dimension.

It is important to note that a signal is not collected during the t_1 period. However, the frequency and amplitude of magnetisation detected during t_2 is modulated by spin evolution that occurs during t_1. This information is transferred to the second spin during the mixing period via spin coupling. To properly see the effect of t_1 on t_2 a number of spectra are recorded with variable values of t_1 . Initially, the value of t_1 is set to zero and the data recorded. t_1 is then set to time Δ_1 and the data recorded again. This is repeated for evolution times $2\Delta_1$, $3\Delta_1$, $4\Delta_1$ and so on (Figure 2.2.3B). These incremental steps may be repeated 100-1000 times for a given experiment, allowing us to record as many FIDs. These will have to be Fourier transformed twice to convert data from being time-dependent (t_1 and t_2) to being frequency dependent (F_1 and F_2). F_1 and F_2 are then plotted as a topographical map with contour lines displaying peak intensity. This map can then be used to derive the interactions between spins of

a system, either from scalar or dipolar coupling.

2.2.6 Three-Dimensional NMR

For small peptides, 2D NMR may provide adequately clear spectra for full system characterisation. However, for full proteins, even relatively small ones of around 5 kDa, there will be a severe degree of overlap in the peaks. This is is particularly prevalent in more complex experiments where multiple J-coupled or NOE peaks are detected. Peak overlap can hinder data analysis and resonance assignment. To circumvent this issue we can resolve the data in a third dimension to separate the overlapped peaks.

The general scheme of a 3D experiment is very similar to that of the 2D experiment except that t_2 is now also an evolution period followed by a second mixing period. This precedes the detection period now labelled as t_3. Fourier transformation of these three time-dependent functions into the frequencies F_1, F_2 and F_3 allows the generation of a 3D spectrum.

It is difficult to predict with any certainty the chemical shifts of nuclei in a protein due to the myriad of factors that influence chemical shift values [132]. Therefore, the network of spin connections is studied to infer which amino acid the spin system relates to. The most commonly used method to achieve this is that of 3D triple-resonance NMR. The technique was first developed during the 1990s and has gained prominent use in the assignment of protein structures [142]. It utilises J-coupling between individual amino acids to show the network of bonds that exist along the protein sequence. Experiments that are used to assign the backbone sequence of the proteins are used in pairs with one experiment showing peaks for the i-1 residues and the corresponding experiment showing peaks for i and i-1 [142–144].

2.2.6.1 Triple Resonance 3D NMR

The HNCO experiment shows the correlation between ^{1}H and ^{15}N spins of residue i with the ^{13}CO spins of residue i-1. Firstly, magnetisation is passed from ^{1}H to ^{15}N, via the $^{1}J_{N,\,H}$ coupling, and then to the ^{13}CO nuclei via the $^{1}J_{N,\,CO}$ coupling (Figure 2.2.4). The experimental pair to the HNCO experiment is the HN(CA)CO experiment. Unlike the HNCO experiment, which only shows correlations to i-1 ^{13}CO spins, the HN(CA)CO shows correlations to i and i-1 ^{13}CO spins. The sequence of magnetisation transfer starts with transfer from ^{1}H to ^{15}N. Magnetisation is then transferred to $^{13}C\alpha$ spins, via the $^{1,2}J_{N,\,CA}$ coupling, and then to the ^{13}CO spins via the $^{1}J_{CA,\,CO}$ coupling (Figure 2.2.4). For both experiments, chemical shift is evolved first on the ^{13}CO nuclei (t_1) and then the ^{15}N nuclei (t_2). Finally, magnetisation is passed back to the ^{1}H nuclei for signal detection (t_3).

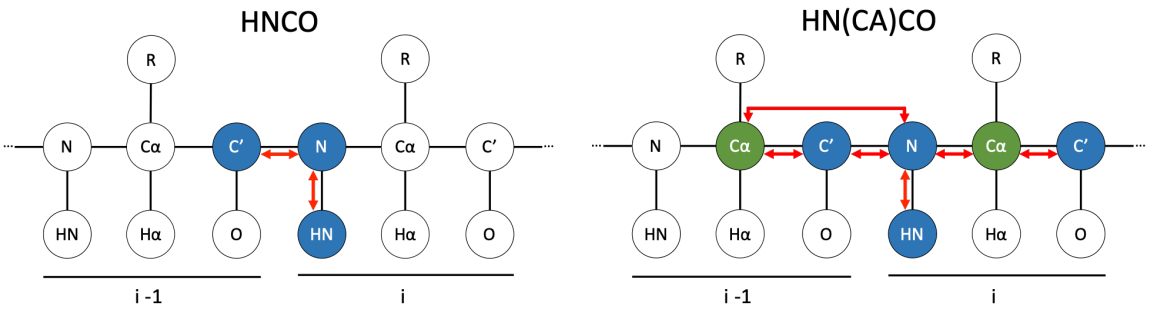


Figure 2.2.4: Schematic of HNCO/HN(CA)CO magnetisation transfer. Blue circles indicate nuclei for which magnetisation is evolved leading to a detectable signal and green circles indicate nuclei for which no evolution is performed. Red arrows indicate the transfer path of magnetisation.

In the HNCACB, and the analogous CBCANH experiment, correlations are shown between the ^{1}H and ^{15}N spins of residue i and the α and β ^{13}C spins of residues i and i-1. The HNCACB is often used over the CBCANH as it offers higher sensitivity for larger proteins with shorter T_2 relaxation times [145]. However, the CBCANH experiment is described here due to the reduced complexity of magnetisation transfer. Magnet-

isation is initially transferred from ^{1}Hα/β to ^{13}Cα/β spins via the $^{1}J_{C,\,H}$ coupling. Magnetisation is evolved on the α and β carbons (t_1) and the then passed from ^{13}Cβ to ^{13}Cα. Depending on whether magnetisation is on residue i or i-1, it is passed to the ^{15}N spin via the $^{1}J_{C,\,N}$ or the $^{2}J_{C,\,N}$ coupling, respectively. ^{15}N chemical shift is then evolved during the t_2 period before magnetisation transfer to the ^{1}H spin for signal detection (t_3) (Figure 2.2.5).

In the CBCA(CO)NH, or HN(CO)CACB, correlations are shown between the ^{1}H and ^{15}N spins of residue i and the α and β ^{13}C spins of the i-1 residue only. Magnetisation transfer is similar but includes transfer from the ^{13}Cα spin to the carbonyl-^{13}C spin via the $^{1}J_{C,\,C}$ coupling (Figure 2.2.5). This limits magnetisation transfer to the i-1 residue only. Chemical shift evolution is the same as it is in the CBCANH experiment.

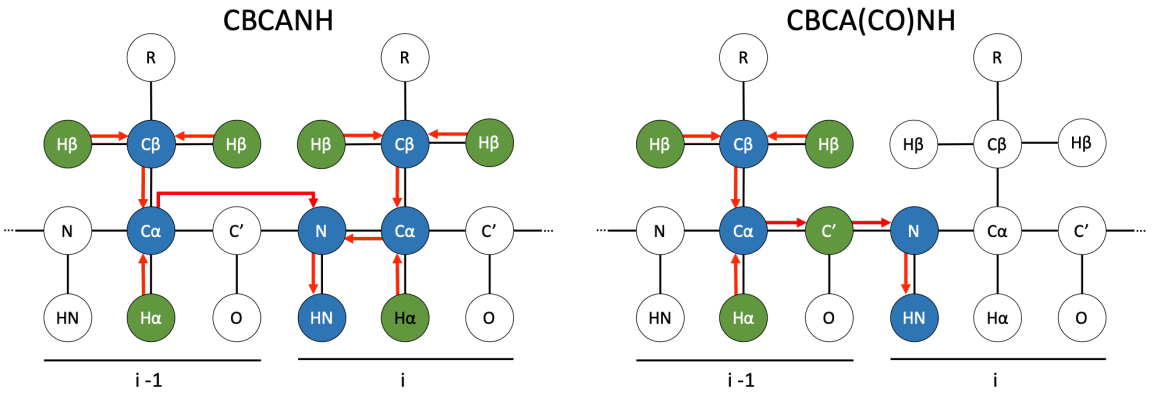


Figure 2.2.5: Schematic of CBCANH/CBCA(CO)NH magnetisation transfer. Blue circles indicate nuclei for which magnetisation is evolved leading to a detectable signal and green circles indicate nuclei for which no evolution is performed. Red arrows indicate the transfer path of magnetisation.

The HNCA experiment correlates the ^{1}H and ^{15}N spins of residue i with the ^{13}Cα spins of residues i and i-1. It is therefore similar to the HNCACB experiment and should not be necessary for backbone assignment. However, if the HNCACB/HN(CO)CACB experiments prove difficult to assign accurately, then the HNCA/HN(CO)CA experiments can be helpful due to the reduced complexity and higher sensitivity of these

spectra. Magnetisation is transferred from ^{1}H to ^{15}N spins via the $^{1}J_{N,H}$ coupling. The $^{1}J_{N,CA}$ coupling is then used to pass magnetisation to the $^{13}C\alpha$ spin and evolved for the t_1 time period. This is followed by magnetisation transfer back to ^{15}N spin and evolution for the t_2 time period. Finally, magnetisation is transferred to the ^{1}H spin for acquisition (t_3) (Figure 2.2.6).

The HN(CO)CA experiment correlates the ^{1}H and ^{15}N spins of residue i with the $^{13}C\alpha$ spins of residue i-1. The selection of inter-residue magnetisation transfer is achieved in the same way as in the CBCA(CO)NH. Magnetisation is transferred from $^{13}C\alpha$ to ^{15}N via the ^{13}CO spins which only occurs for i-1 α-carbons (Figure 2.2.6). The evolution periods are the same as in the HNCA experiment.

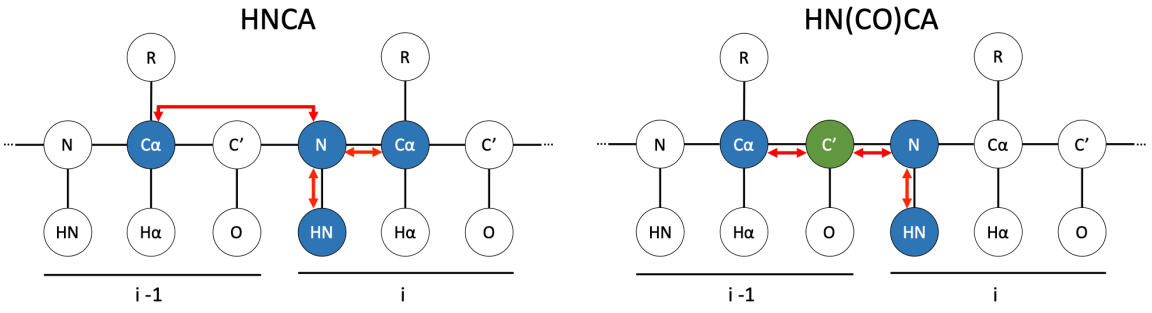


Figure 2.2.6: Schematic of HNCA/HN(CO)CA magnetisation transfer. Blue circles indicate nuclei for which magnetisation is evolved leading to a detectable signal and green circles indicate nuclei for which no evolution is performed. Red arrows indicate the transfer path of magnetisation.

2.2.7 COSY and TOCSY Experiments

COSY and TOSCY experiments can be recorded in two or three dimensions. For small peptides the separation in two dimensions may be sufficient to allow full peak assignment. However, proteins above around 3 kDa the three-dimensional experiments are used to reduce peak clustering.

The correlation spectroscopy (COSY) experiment was the first 2D experiment to be

developed [140]. The 2D version of the experiment is homonuclear with ^{1}H frequencies in both x and y dimensions. The 3D experiment, which is used for most proteins, is a double-resonance experiment with the ^{13}C frequencies in the z axis.

The spectra show two types of peaks. Those that appear along the diagonal (F_1= F_2) represent the peaks of the 1D proton spectrum. Cross-peaks, which have different F_1 and F_2 frequencies, arise from magnetisation transfer between J-coupled spins. It is the occurrence of these cross-peaks that allows correlation between the two diagonals that contribute to the cross-peaks. A schematic of a COSY spectrum for a simple three spin system is shown in Figure 2.2.7.

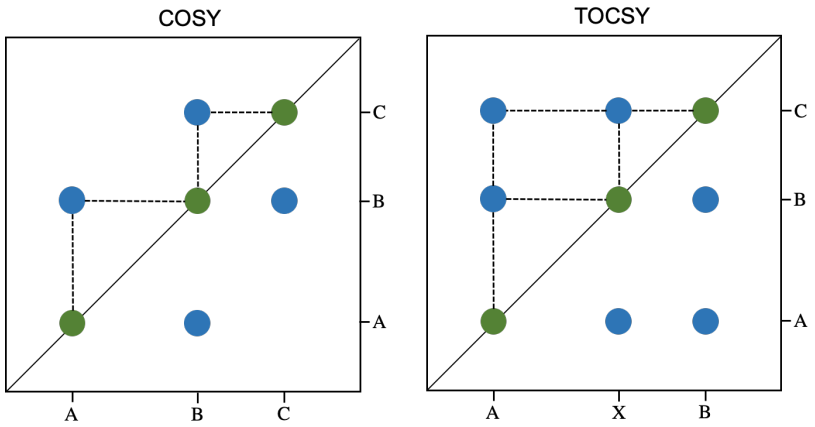


Figure 2.2.7: Schematic of a three-spin system COSY and TOCSY spectrum. Diagonal peaks (green) show the $F_1$and F_2 frequencies for a given spin in the absence of coupling. The cross-peaks (blue) arise due to the J-coupling between two spins. For example, the F_1 frequency from spin A and the F_2 frequency from spin B. The COSY spectrum shows only directly bond partners. However, The TOCSY experiment shows J-coupling over the length of several covalent bonds.

In the HCCH-COSY experiment correlations are made between neighbouring side-chain ^{1}H and attached ^{13}C resonances [142, 146, 147]. If we are to take a Hγ resonance as an example, then magnetisation is first generated on the ^{1}Hγ resonance and evolved for the t_1 time period. Magnetisation is then transferred to the ^{13}C γ-carbon resonance via the resonance $^{1}J_{C,\,H}$ coupling and evolved for the t_2 time period. The $^{1}J_{C,\,C}$ coupling is then used to pass magnetisation to the neighbouring ^{13}C resonances (Cβ).

Magnetisation is then passed to the $^{1}H\beta$ resonance for carbon-decoupled acquisition (t_3).

The magnetisation transfer is very similar for the total correlation spectroscopy (HCCH-TOCSY) experiment. However, there is further mixing to allow magnetisation to pass to all the side-chain ^{13}C resonances, utilising the $^{1}J_{C,\,C}$ coupling. This leads to chemical shift evolution on Cα, Cβ and Cγcarbons during the t_2 time period. Similarly, during the t_3 time period, a signal is acquired from the Hα,Hβ and Hγproton resonances.

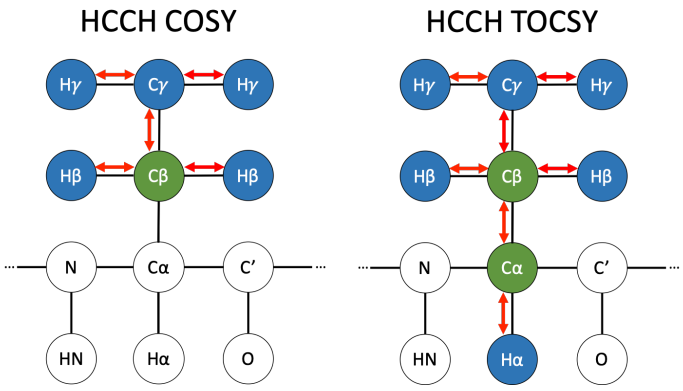


Figure 2.2.8: Schematic of COSY/TOCSY magnetisation transfer. Blue circles indicate nuclei for which magnetisation is evolved leading to a detectable signal and green circles indicate nuclei for which no evolution is performed. Red arrows indicate the transfer path of magnetisation.

2.2.8 NOESY Experiments

The experiments described above allow the assignment of chemical shifts for most backbone and side-chain resonances. However, they cannot provide information about the tertiary structure of the protein. To generate a 3D structural model of a protein the highest possible number of distance restraints should be generated. The most powerful and commonly used method to achieve this is by using a NOESY (Nuclear Overhauser Effect SpectroscopY) experiment.

In the NOESY-HSQC experiment magnetisation is transferred between ^{1}H resonances using the NOE and evolved for the t_1 time period. Magnetisation is then transferred to the heteronuclei, either ^{15}N or ^{13}C, via the $^{1}J_{X,H}$ coupling and evolved for the t_2 time period (Figure 2.2.9). Finally, magnetisation is transferred back to the ^{1}H resonance via the same J-coupling and the signal recorded (t_2). This results in a 3D spectrum that has ^{1}H-X diagonal peaks, arising from J-coupling and ^{1}H-^{1}H cross-peaks arising from the NOE. The intensity of the peak corresponds to the distance between the coupled spins which means the experiment can be used to determine the tertiary structure of the protein.

The process can become complicated if spin X has two dipolar coupled spins, A and B, with a similar chemical shift, leading to overlapping peaks in the resulting spectrum. In addition, incomplete resonance assignment can lead to missing or incorrect assignment. For these reasons, peaks in the NOESY spectrum are not given definitive assignments but are ambiguously assigned to a number of possible resonances.

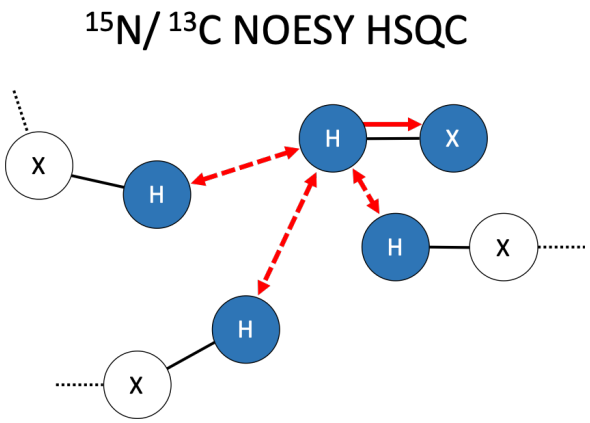


Figure 2.2.9: Schematic of NOESY magnetisation transfer. The X nuclei J-coupled to the H nuclei can be either carbon or nitrogen giving rise to the ^{13}C-NOESY or ^{15}N-NOESY, respectively.

2.3 NMR Spectroscopy to Derive Structure and Function

2.3.1 Generation of protein structural models

By assigning the resonances of a protein, using the experiments described in this chapter, it is possible to generate a structural model for the protein. There are two main constituents required to do this. Firstly, ambiguous assignment of ^{15}N and ^{13}C NOESY are used to determine the inter-atomic distance between nuclei of the system. The greatest possible number of these are used to improve the accuracy of the model. However, overlapping or broadening of peaks means that some are discarded during distance calculation and only the restraints that are consistent with a compact folded structure are included [132].

In addition to NOEs, shift-based dihedral angle restraints can be used to improve the accuracy of the structural model [148]. Indeed, several methods for structure determination use dihedral angle predicted fragments to construct initial models [149, 150]. The DANGLE (Dihedral ANgles from Global Likelihood Estimates) algorithm uses Bayesian inference to estimate the likelihood of conformations throughout Ramachandran space [149]. The algorithm uses the expected population distributions for different amino acid types to generate dihedral restraints for each residue in the protein

The dihedral restraints and NOE peaks are used to generate structural models via an automated process, implemented by programs such as ARIA or CYANA [151, 152]. The software takes the spectral information and uses it to produce initial structural models of the protein. Assignments that conflict with the structural models are discarded and new models created. Several iterations of this process are performed to further refine the models.

The researcher will have to filter through restraints that violate the structural models. The erroneous restraints may arise due to incomplete and/or incorrect peak assignment or due to peak overlap that means unambiguous assignment is not possible [153]. Such errors can be reviewed and either removed from the calculation or corrected, if possible.

It is important to be aware that structures produced do not represent the 'true' structure, but a model that aligns with experimental observations [132]. Therefore, there are two main classes of structural statistics that are used to gauge the quality of the model. Knowledge-based metrics measure model conformance to expected values via comparison to databases of experimentally determined structures. The metrics include the distribution of dihedral angles, bond length, the geometry of hydrogen bonds and atomic packing [154]. Model versus data metrics measure the alignment of structural models to the experimental data. It is an assessment of the data used to derive the model and includes measures such as assignment completeness and the violation of restraints. Root mean square deviation (RMSD) for backbone or all heavy atoms is also used as a measure of structure precision [154].

2.3.2 Probing Protein Structural Dynamics by NMR

Of the techniques available for studying proteins at the atomic level, only NMR spectroscopy provides temporal information down to the ps timescale. As described in Section 2.1.5 molecular motion is the primary source of local magnetic fields that drive spin relaxation. As such, the rate of relaxation can be measured to infer information about the dynamic state of the protein molecule.

There are many possible mechanisms that contribute to the rate of relaxation. However, when studying relaxation for proteins the primary mechanism observed is that of dipole-dipole (DD) relaxation. The local magnetic field fluctuations that arise

from dipolar coupled nuclei causes the magnetisation vector to relax, as described in Equation 2.3.1.

$$R_{1(DD)} = \frac{1}{T_{1(DD)}} = \gamma_X^2 \gamma_H^2 \left(\frac{h}{2\pi}\right)^2 \tau_c \sum \frac{1}{r_{HX}^6} \tag{2.3.1}$$

Where, $\gamma_X \gamma_H$ is the gyromagnetic ratio of the coupled spins, h is Plank's constant, τ_c is correlation time and r_{HX} is the distance between the two nuclei. Given that the mechanism is dependent on the gyromagnetic ratio of the interacting nuclei, protons are the most effective at causing DD relaxation. This equation also tells us that nuclei located in a protein core, experiencing many dipolar couplings, will relax faster than nuclei located in flexible disordered regions where very few dipolar couplings exist. Correlation time can also vary within a single molecule with large globular regions taking much longer to make one complete rotation relative to flexible coil regions.

T_1 relaxation is measured using inversion recovery. This involves the use of a hard pulse followed by a variable delay period (τ) to allow for spin-lattice relaxation. Proton saturation is applied to suppress relaxation via other mechanisms. The measurement of T_2 relaxation is achieved using the Carr-Purcell-Meiboom-Gill sequence (CPMG) pulse pattern. This pattern is based on the spin-echo pulse pattern with variable τ times used to allow for spin-spin relaxation. Both of these techniques can be incorporated into HSQC experiments to measure the relaxation of each NH nuclei in a given protein sample.

The NOE effect, described in Section 2.1.6.2, is dependant on the correlation time of the coupled nuclei. It is theoretically possible therefore, to determine the correlation time from homonuclear NOESY experiments. However, this is very difficult in practice because the true distance between the dipolar coupled nuclei is often not known and multiple similar couplings may be occurring simultaneously [155].

To address this difficulty heteronuclear NOE experiments can be used which meas-

ure one proton and a covalently linked heteronucleus, often the backbone ^{1}H-^{15}N. Given that the bond distance is well defined this makes analysis easier and more accurate. Bond vectors that undergo molecular motion that is faster than the average correlation time of the entire protein (pico- to nanosecond time scale) have a reduced NOE intensity relative to the average NOE observed for most bond vectors in the protein. Fewer NOEs are often observed at the protein N- and C-terminus and in regions of increased flexibility.

Values for the heteronuclear-NOEs are calculated from the ratio of peak intensities according to Equation 2.3.2.

$$NOE = (I_{saturated}/I_{equilibrium}) - 1 \tag{2.3.2}$$

Where, $I_{saturated}$ refers to the peak intensity in the spectrum where proton saturation is applied during the preparation stage, and $I_{equilibrium}$ refers to peak intensity where the proton spins are left at equilibrium prior to the ^{15}N excitation pulse. By saturating the proton signal cross-relaxation occurs via the NOE causing a loss of magnetisation. This effect can therefore be used as a proxy to study correlation time and local differences in molecular motion.

Chapter 3

Materials and Methods

3.1 Strains and Plasmids

Bacterial strains and plasmids used in this study are listed in Appendix A.

3.2 Oligonucleotides

IDT OligoAnalyzer Tool was used to check oligonucleotide melting temperatures, GC content and dimerisation probability. Oligonucleotides were selected to have a GC content as close to 50% as possible. Sequences were also selected to have the lowest possible probability to form either self- or hereto-dimers. Where possible the last five bases at the 3' end of the oligonucleotide was designed to have two or three G or C bases to promoter more specific 3' annealing. The majority of oligonucleotides were designed to be 18-22 bp in length. For Gibson assembly oligonucleotides were deigned to possess at least 20 bp of overlap with adjacent fragments. Therefore, these oligonucleotides were significantly longer. All oligonucleotides were synthesised and delivered by IDT.

All oligonucleotides used in this study are listed in Appendix A.

3.3 preparation of Growth Media

Granulated Luria Agar was used to make agar plates for growth of *E. coli* cells. 37 g of granulated powder was added to 1 L of ddH_2O and autoclaved for 15 minutes. The solution was cooled to 50 °C and the appropriate antibiotics added. 20 mL aliquots were made into 90 mm petri dishes in a laminar flow cabinet to maintain sterility. Plates were stored at 4 °C prior to use.

E. coli cells were grown in Miller's Luria broth (LB) or 2XYT broth for the production of unlabelled proteins. 25 g granulated LB powder or 31 g granulated 2XYT powder was added to 1 L of ddH_2O and the pH adjusted to 7.2. Media solutions were autoclaved for 15 minutes and stored at room temperature for no more than one week. Antibiotics were added to the media just prior to use.

Table 3.3.1: Recipe for M9 minimal media.

Component	Amount (per Litre)
Na_2HPO_4	6.0 g
KH_2PO_4	3.0 g
NaCl	0.5 g
Glucose ($^{13}C_6$ glucose for carbon labelling)	2.0 g
NH_4Cl ($^{15}NH_4Cl$ for nitrogen labelling)	1.0 g
1 M $MgSO_4$	2.0 mL
1 M $CaCl_2$	10 μL
100 mM $FeSO_4$	100 μL
Trace Element Solution (100X; ThermoFisher Scientific)	10 mL
Micronutrient solution (1000X)*	1 mL

* (3 μM Ammonium molybdate; 400 μM H_3BO_4; 30 μM $CoCl_2$; 10 μM $CuSO_4$; 80 μM $MnCl_2$; 10 μM $ZnSO_4$).

M9 minimal media (Table 3.3.1) was used to grow *E. coli* cells producing ^{13}C and/or ^{15}N labelled proteins. Na_2HPO_4, KH_2PO_4 and NaCl were dissolved in 0.8 L ddH_2O and adjusted to pH 7.4. The solution was then made up to 1 L and autoclaved for 15 minutes. Prior to use all remaining reagents in Table 3.3.1 were added

to the solution along with the appropriate antibiotics. D-glucose-$^{13}C_6$ and $^{15}NH_4Cl$ (Cambridge Isotopes) were used, where appropriate for isotopic labelling of the recombinantly expressed protein.

Blood agar plates were used for plate growth of *P. gingivalis*. 40 g blood agar base (Oxoid) was added to 950 mL ddH_2O and autoclaved for 15 minutes. The solution was cooled to 50 °C followed by the addition 50 mL sterile defibrinated horse blood (TCS Biosciences) and appropriate antibiotics.

Liquid culturing of *P. gingivalis* was performed using brain heart infusion (BHI) broth (Oxoid). 37 g BHI powder was dissolved in 1 L ddH_2O and autoclaved for 15 minutes. Prior to use the media was supplemented with 5 µg/ mL haemin (BHI-H) and appropriate antibiotics. BHI-H was equilibrated to anaerobic conditions for at least 24 hours prior to use.

The antibiotic stocks used in bacterial culturing were all prepared at 1000× concentration and filter sterilised using a 0.22 µm filter syringe. Sterilised antibiotic stocks were stored at -20 °C to prevent degradation. The stock concentration and solvent for all antibiotics used in this study are listed in Table 3.3.2. Where possible, ethanol was used as a solvent to prevent stocks from freezing. All chemicals used in the preparation of bacterial cultures and antibiotics were provided by Melford Laboratories Ltd.

Table 3.3.2: Antibiotic stock specifications.

Antibiotic	Abbreviation	Stock Concentration	Solvent
Ampicillin	Amp	100 mg mL^{-1}	50% Ethanol
Chloramphenicol	Cam	30 mg mL^{-1}	50% Ethanol
Erythromycin	Erm	50 mg mL^{-1}	95% Ethanol
Kanamycin	Kan	50 mg mL^{-1}	ddH_2O
Tetracycline	Tet	12.5 mg mL^{-1}	50% Ethanol

3.4 Bioinformatic Analysis

Protein sequences were submitted to the ExPasy ProtParam server [156, 157] was used to calculate the protein isoelectric point (pI), protein molecular weight, extinction coefficient and grand average of hydropathy [158] (GRAVY) score for the protein of interest. Additional sequences, such as tags and protease cleavage sites, were included in parameter predictions where appropriate.

Protein signal sequences were identified using the online SignalP 5.0 server [159]. Search queries were limited to the Gram-negative bacteria database. The prediction of transmembrane beta-strands was achieved using the PRED-TMBB web-server [160, 161].

NCBI BlastP was used to identify similar protein sequences for multiple sequence alignments [162]. The default settings were using with an except threshold of 0.05 and the maximum number of target sequences adjusted to 500. The online MUSCLE server was used for multiple-sequence alignments of less than 50 sequences, with default settings [163]. For large multiple-sequence alignments (more than 50 sequences) the online Clustal Omega server was used, with default settings[164]. FASTA output format was selected for visualisation of sequence alignments. The FASTA output files from the multiple sequence files were uploaded to the WebLogo server [165], using default settings, to generate consensus sequence images. The multiple sequence alignment results were uploaded to the online ConSurf server [166, 167], along with the appropriate structural model, to produce structural models coloured by sequence conservation. The default settings for the ConSurf server were used.

The online PHYRE^2 server was used for the prediction of protein tertiary structure based on protein sequence [168]. Signal peptide sequences, for transport through the Sec system, were excluded from model calculation. Intensive modelling mode was selected for model calculation.

3.5 Molecular Cloning

3.5.1 Polymerase Chain Reaction

DNA fragments for use in molecular cloning were amplified via polymerase chain reaction (PCR). Q5 High-Fidelity DNA Polymerase (NEB) was used in all PCR reactions following the 25 µl reaction protocol according to manufacturers instructions (Table 3.5.1). The design of oligonucleotides used as primers is described in Section 3.2. Reactions were performed in a TC-312 thermo-cycler (Techne) using the settings shown in Table 3.5.2.

Table 3.5.1: PCR reaction components.

Component	Volume
5×Q5 reaction buffer	5 µl
10 mM dNTPs	0.5 µl
10 µM forward primer	1.25 µl
10 µM reverse primer	1.25 µl
Template DNA	variable*
Q5 high-fidelity DNA polymerase	0.25 µl
Nuclease-free water	to 25 µl

* 50-100 ng of genomic DNA template and 10-50 ng of plasmid DNA was used per reaction.

3.5.2 Restriction Enzyme Digestion

All restriction enzymes used in this study were provided by NEB and used according to manufacturers instructions. All digestions performed using CutSmart buffer (NEB). Reactions were incubated at 37 °C for 1 hour, immediately followed by agarose gel electrophoresis.

Table 3.5.2: PCR Thermocycler Settings.

Temperature (°C)	Time (seconds)	Cycles
98	30	1
98 x* 72	20 20 30 (per 1000bp)	x30
72	180	1
4	Hold	-

* Annealing temperatures used were 5°C lower than the lowest Tm of the primer pairs used.

3.5.3 Agarose Gel Electrophoresis

DNA samples were separated via agarose gel electrophoresis using 1-1.5% agarose gels prepared with TAE buffer (40 mM Tris pH 7.6, 20 mM acetic acid, 1 mM EDTA) and SYBR safe stain (ThermoFisher Scientific). Gels were run using a PowerPac™ Basic Power Supply (BioRad) at 90 V for 30 minutes. 2-log DNA ladder (NEB) was used as a marker to determine band size.

3.5.4 DNA Purification

3.5.4.1 Genomic DNA Purification

P. gingivalis W50 genomic DNA was extracted for use as a template in PCR. A 10 mL culture grown overnight at 37 °C was used for each purification. The DNeasy UltraClean Microbial kit (Qiagen) was used according manufacturer instructions. Genomic DNA samples were stored at -20 °C.

3.5.4.2 Plasmid DNA Purification

The Monarch® Plasmid Miniprep Kit (NEB) was used for the extraction of plasmid DNA from *E. coli*. A total volume of 3 mL bacterial culture, taken from a 10 mL

overnight culture, was used for each miniprep. Purifications were performed according to manufacturers instructions. Samples were eluted in 30 µl Nuclease-free water. Plasmid DNA samples were stored at -20 °C.

3.5.4.3 PCR Product Purification

PCR amplified products were purified using the Monarch® PCR & DNA Cleanup Kit (NEB) according to manufacturers instructions. Samples were eluted using 10-15 µl Nuclease-free water. This purification was used to remove reagents, unwanted buffers and impurities from the sample prior to restriction endonuclease digestion or ligation.

3.5.4.4 Agarose Gel Band Extraction and Purification

DNA bands of the appropriate size were excised from agarose gels using a sterile scalpel. The sample was then purified using a GeneJET Gel Extraction Kit (ThermoFisher Scientific) or a QIAquick Gel Extraction Kit (Qiagen) according to manufacturer instructions.

3.5.5 Plasmid Assembly

3.5.5.1 Ligation Independent Cloning of His-RgpB-CTD

The pET-46 Ek/LIC Vector Kit (Novagen) was used to perform ligation independent cloning (LIC) for the His-RgpB-CTD construct. The LIC compatible purified PCR product of the RgpB-CTD gene fragment was treated by T4 DNA Polymerase (ThermoFisher Scientific) and annealed into the pET-46 vector according to manufacturers instructions. NovaBlue GigaSingles™ Competent Cells (Novagen), provided in the kit, were used for heat-shock transformation of the assembled plasmid. Insert sequences were confirmed by Sanger sequencing (Eurofins).

3.5.5.2 Gibson Assembly of pTIO-AI Constructs

pTIO-AI constructs were assembled using the NEBuilder® kit (NEB). pTIO-I vector was cut using BamHI and NcoI and the digested plasmid purified from agarose gel extraction. PCR fragments were amplified using primers with >20 bp overhangs and purified prior to use in assembly. 0.05 pMol of each fragment was used for each reaction at an 1:1 molar ratio. DNA fragment molarity was calculated using the online NEBioCalculator™ tool. Fragments were ligated using the NEBuilder master mix, according to manufacturers instructions, allowing 1 hour for ligation reactions to proceed. Insert sequences were confirmed by Sanger sequencing (Eurofins).

3.5.5.3 Molecular Cloning of PorV Construct

The pET28(b)-OSS plasmid was provided by Sunjun Wang (KCL) for cloning of the PorV construct. The vector, and corresponding PCR fragment, were cut using HindIII and NotI. The pASK-IBA2C vector, and corresponding PorV PCR product, were cut using SacI and XhoI. The digested products were separated by agarose gel electrophoresis and the bands purified (Section 3.5.4.4).

Plasmid ligations were performed using T4 DNA ligase (NEB) according to manufacturer instructions. Molar ratios of 1:3, 1:5 and 1:10 of vector to insert were screened. Room temperature incubation for 10 minutes was used, followed by heat inactivation at 65 °C for 10 minutes. Ligation samples were used immediately for bacterial transformation and remaining sample stored at -20 °C.

3.5.6 Bacterial Transformation

3.5.6.1 Heat-Shock

20 µL chemically competent cells were incubated on ice with plasmid for 20 minutes. Cells were then heat shocked for 30 seconds using a 42 °C water bath then incubated

on ice for 2 minutes. Outgrowth of transformed cells was achieved using 400 µl LB with 1 hour of 37 °C incubation. 50-200 µl was then taken and plated on LB agar plates with appropriate antibiotics for selection.

3.5.6.2 Electroporation

Electrocompetent *P. gingivalis* cells were prepared using a 6 hour 30 mL culture harvested by centrifugation at 10,000 × g for 10 minutes. The cell pellet was resuspended in 1 mL ice-cold sterile electroporation buffer (EPB; 10% glycerol-1mM $MgCl_2$). Cells were pelleted again and washed two more times with 1 mL EPB and again suspended in 1 mL EPB. Electrocompetent *P. gingivalis* cells were then used immediately for electroporation.

100 ng of plasmid DNA was mixed with the 200 µl aliquots of the cell suspension. The mixture was then electroporated (2500V, 25 μF capacitance and 200Ω resistance) in a 2 mm electroporation cuvette. 1 mL of room temperature BHI-H was immediately added and cells allowed to recover for 16 hours at 37 °C under anaerobic conditions. Cells were then plated onto blood agar plates with appropriate antibiotics.

3.6 Protein Techniques

3.6.1 Expression and Purification of Recombinant CTD

3.6.1.1 Protein Expression

pET46-His-RgpB-CTD was transformed in to BL21 (DE3) cells using heat-shock transformation. The transformed cells were grown on Luria agar plates overnight. 3-5 colonies were taken from the plate and used to inoculate 10 mL LB cultures in 30 mL universal containers which were grown at 37 °C for 4 hours with 240 rpm shaking. A single 10 mL culture was then used to inoculate 0.5 L of LB or M9 minimal

media, for the preparation of isotopically labelled protein, in a 2.5 L full-baffle flask. These cultures were grown, under the same conditions, to an optical density at 600 nm (OD_{600}), of between 0.6 - 0.8. Absorbance was measured for bacterial cultures using UV spectroscopy after blanking the spectrometer with LB media. After achieving the correct density, culturing temperature was decreased to 18 °C and protein expression induced by the addition of 100 µM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were then incubated for a further 16 hours to allow for protein expression. Ampicillin was used for positive selection at all stages of growth and expression.

3.6.1.2 Ni-NTA Chromatography

Following protein expression cells were harvested by centrifugation at $5000 \times \mathbf{g}$ for 15 minutes at 4 °C. The cell pellet was resuspended in the ice-chilled lysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5 mM $MgCl_2$, 6 µg/ mL lysozyme, 1 µg/ mL DNase I (NEB) and 1 mM PMSF) at approximately 5 g mL^{-1} of wet cell weight. Samples were then sonicated, using a Sonics® Vibra Cell VC130 sonicator, for 3 minutes at 70% power with 3 second pulses. Sonicated samples were centrifuged at $25,000 \times \mathbf{g}$ for 30 minutes at 4 °C to remove cellular debris.

The cell lysate was loaded onto 5 mL HF HisTrap column (GE Healthcare) pre-equilibrated with Ni-NTA wash buffer (20mM Tris-HCl, pH 8.0, 300 mM NaCl, 10 mM imidazole). The column was washed using an AKTA Pure (GE Healthcare) until A280 absorbance reached a steady baseline. The sample was then eluted in 2 mL fractions using a linear gradient from 100% Ni-NTA wash buffer to 100% Ni-NTA elution buffer (20mM Tris-HCl, pH 8.0, 300 mM NaCl, 300 mM imidazole). Elution was continued until a steady baseline was reached.

3.6.1.3 Enterokinase Cleavage

The Ni-NTA purified sample was first transferred to enterokinase cleavage buffer (20 mM Tris-HCl, pH 8.0, 50 mM NaCl, 2 mM $CaCl_2$) using SnakeSkin™ Dialysis Tubing, 3.5K MWCO (ThermoFisher Scientific). Samples were dialysed using an excess of 500× for external dialysis buffer and left for 16 hours at 4 °C with gentle mixing. Protein concentration was then adjusted to 2 mg mL^{-1} and 16 units of enterokinase light chain (NEB) added for every 10 mL of sample. The reaction was allowed to proceed for 16 hours at 25 °C with gentle mixing.

The cleaved sample was then immediately processed by reverse Ni-NTA chromatography to remove the enterokinase and cleaved His-tag peptide. Reverse Ni-NTA was performed using the same method as above (Section 3.6.1.2). However, lysate flow-through and five column volumes of washing flow-through was taken as the enterokinase cleaved protein sample. Five column volumes of elution buffer was used for a single elution step to collect remaining bound protein.

3.6.1.4 Size Exclusion Chromatography

Following HisTrap or reverse Ni-NTA purification samples were further purified using size exclusion chromatography. This was performed using an AKTA PURE (GE Healthcare) with a Superdex 16/600 S75 column (GE Healthcare). The column was first equilibrated one column volume of buffer containing 20 mM $NaPO_4$, pH 6.0 and 100 mM NaCl. Prior to loading the protein sample was concentrated to ~1 mL and injected onto the column using a 2 mL injection loop. The sample was then separated using a flow rate of 1 mL $min^{-1}$with 2 mL fraction size collection. 20 µl samples were taken from fractions corresponding to major elution peaks for sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis. Fractions corresponding to the correct elution peak were pooled and concentrated using a Vivaspin® 20 (5 kDa

MWCO) centrifugal concentrator (Sartorius) at 4000 × **g** and 4 °C until the desired protein concentration was achieved.

3.6.2 Expression and Purification of Recombinant PorV

3.6.2.1 Protein Expression Trials

Prior to large-scale expression of recombinant PorV, small scale protein expression trials were performed. The relevant plasmids (pASK-IBA2C-PorV, pET28-Oss-His-PorV and pET28-Oss-PorV-His) and were transformed into BL21 (DE3), Lemo21 and C41 cells by heat-shock transformation. However, given that Lemo21 is optimised for T7 promoter plasmids, the pASK-IBA2C-PorV, which contains a Tet promoter, was not transformed into Lemo21. The transformants were grown overnight on Luria agar plates at 37 °C with appropriate antibiotic selection. Single colonies were then transferred to 10 mL LB in 30 mL universal containers containing the appropriate antibiotic. In addition, Lemo21 cultures were supplemented with 0, 100, 250, 500, 750, 1000 and 2000 µM L-Rhamnose. Cultures were grown at 37 °C with 240 rpm shaking to an OD_{600} of ~0.4. Cells carrying pET28-Oss-His-PorV and pET28-Oss-PorV-His were then induced with 0 or 0.4 mM IPTG. Cells carrying pASK-IBA2C-PorV with induced with 0, 1 or 2 µg mL^{-1} anhydrotetracycline (IBA Lifesciences). Cells were grown post-induction at 18 °C for 16 hours with 240 rpm shaking.

0.5 mL samples from each culture were taken and centrifuged at 13, 000 × **g** for 1 minute. The cell pellet was resuspended in 50 µl of 1×BugBuster® Protein Extraction Reagent (ThermoFischer Scientific), prepared from a 10× stock solution diluted in phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and KH_2PO_4, pH 7.4). Cells were incubated at room temperature for 15 minutes with gentle mixing and then centrifuged again at 16, 000 × **g** for 20 min at 4°C to remove insoluble material. The supernatant was then mixed with 50 µl of 2×sodium dodecyl

sulfate (SDS)-loading buffer. Additional 0.5 mL samples were taken from each culture and resuspended in 100 µl 1×SDS-loading buffer. Samples were then analysed by SDS-PAGE and Western blot.

pET28-Oss-His-PorV was selected for additional screening of different IPTG concentrations, OD_{600} for induction and media composition. The method for screening was adapted from the above, using cells grown to an OD_{600} of 0.4, 0.8 and 1.2 in LB media, and 0.8, 1.2 and 1.6 in 2XYT media. Cells were then induced with 0.4 mM IPTG and treated the same as above. 2XYT with induction at $OD_{600}$of 0.8 was selected as the best condition for growth and the screen repeated using a range of IPTG concentrations (0, 50, 100 200, 400 and 600 µM). Expression conditions providing the highest protein yield was selected for large-scale protein expression.

3.6.2.2 Large-scale Protein Expression

3-5 colonies from pET28-Oss-PorV-His in BL21 (DE3) were taken from transformation plates (see above) for inoculation of 10 mL LB cultures in 30 mL universal containers. These were grown at 37 °C for 4 hours with 240 rpm shaking. A single 10 mL culture was then used to inoculate 0.5 L of 2XYT. These cultures were grown, under the same conditions, to an optical density at 600 nm (OD_{600}), of between 0.8 and expression induced by addition of 100 µM IPTG. The temperature was then decreased to 18 °C and the cultures incubated for a further 16 hours to allow for protein expression. Kanamycin was used for positive selection at all stages of growth and expression.

3.6.2.3 Solubilisation Trials

For PorV solubilisation efficiency was tested using the solubilisation trial protocol adapted from Morrison *et al.* [169]. Purified membranes containing the target protein was suspended in buffer containing 20 mM Tris, pH 8 and 200 mM NaCl. The buffer

was supplemented with 2% octyl-β-D-glucoside (OG; Sigma Aldrich), 2% n-dodecyl-β-D-maltopyranoside (DDM; Sigma Aldrich), 1% Lauryldimethylamine oxide (LDAO; Sigma Aldrich), 1% Tetraethylene glycol monooctyl ether (C_8E_4; Sigma Aldrich) and 2.5% styrene maleic anhydride (SMA2000; gifted from Dr. Alan Goddard). The suspension was left to slowly mix at room temperature on a rolling-mixer for one hour. Samples were centrifuged at $100,000 \times \mathbf{g}$, 4 °C for 30 minutes and the supernatant retained as the solubilised protein fraction. The pellet was resuspended in buffer containing 20 mM Tris, pH 8, 200 mM NaCl and 2% SDS.

Soluble vs insoluble fractions for each solubilising agent were analysed by western blot in triplicate. Band intensities were calculated using ImageLab (Biorad) and average band intensities for soluble vs insoluble material used to calculate solubilisation percentage.

3.6.2.4 Membrane extraction

For extraction of bacterial membranes containing folded PorV protein large-scale cultures were collected after 16 hours of protein expression by centrifugation at $5000 \times \mathbf{g}$ for 15 minutes at 4 °C. The pellet was resuspended in the ice-cold lysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5 mM $MgCl_2$, 6 µg/ mL lysozyme, 1 µg/ mL DNase I and 1×cOmplete™ EDTA-free protease inhibitor cocktail (Sigma Aldrich)) at approximately 5 g mL^{-1} of wet cell weight. Samples were then sonicated for 3 minutes at 70% power with 3 second pulses. Sonicated samples were centrifuged at $20,000 \times \mathbf{g}$ for 30 minutes at 4 °C to remove cellular debris. The supernatant was collected and centrifuged at $100,000 \times \mathbf{g}$, 4 °C for 1 hour to pellet bacterial membranes. The pellet was resuspended in buffer containing 20 mM Tris, pH 8, 200 mM NaCl and 1% LDAO. The suspension was left to mix for 45 minutes at room temperature on a rolling mixer. The suspension was centrifuged again at $100,000 \times \mathbf{g}$ for 1 hour at

4 °C to remove insoluble material. The supernatant was retained as the solubilised membrane fraction.

3.6.2.5 Ni-NTA Chromatography

Ni-NTA Chromatography for His-PorV constructs were purified using a poly-prep gravity flow column containing 1 mL Ni-NTA agarose resin (ThermoFisher Scientific). The column was equilibrated with 10 column volumes of Ni-NTA wash buffer and the solubilised membrane sample loaded. The column was washed using 20 column volumes of Ni-NTA wash buffer. A series 2 mL elution factions were collected using Ni-NTA elution buffer. All Ni-NTA buffers were supplemented with 0.1% LDAO to maintain protein solubility.

3.6.2.6 Size Exclusion Chromatography

Size exclusion of Ni-NTA purified PorV was performed using the same technique as described in Section 3.6.1.4 with the following modifications. A Superdex 16/600 S200 column (GE Healthcare) was used following equilibration with buffer containing 20 mM Tris, pH 8, 200 mM NaCl and 0.1% LDAO. Prior to loading the protein sample was concentrated to ~0.5 mL and injected onto the column using a 1 mL injection loop. Pooled fractions were concentrated using a Vivaspin® 20 (30 kDa MWCO) centrifugal concentrator (Sartorius) at 4000 × **g** and 4 °C until the desired protein concentration was achieved.

3.6.3 Oligomerisation Analysis

3.6.3.1 Glutaraldehyde Cross-linking of RgpB-CTD

Cross-linking of His-RgpB-CTD was performed according to de Diego *et al.,* [170] using protein concentrations of 0.1, 0.5 and 1.0 mg/mL with glutaraldehyde at 0.01% in PBS.

Reactions were performed at at 37 °C for 5 minutes followed by reaction quenching by addition of Tris to a final concentration of 100 mM. Cross-linking quenching was performed at room temperature for 15 minutes. Sample was mixed SDS-PAGE loading buffer and heated to 70 °C for 30 minutes to prevent cross-link breakage.

3.6.3.2 DSP Cross-linking of PorV

A 50 mM stock of dithiobis succinimidyl propionate (DSP) was prepared using 100% dimethyl sulfoxide (DMSO). Cross-linking reactions for His-PorV were performed using a series of 20 µl reactions containing 0.5 mg mL^{-1}, ~12 µM, His-PorV in PBS supplemented with 0.1% LDAO. DSP was added to the suspension to a final concentration of 120 µM. Samples were incubated at room temperature for a time series of 0-60 minutes, with 10 minute intervals. 1 M Tris, pH 7.4 was added to a final concentration of 100 mM and incubated for a further 15 minutes at room temperature to quench the reaction.

3.6.4 Protein Quantification & Qualification

3.6.4.1 Determination of protein concentration

Protein sample concentration was quantified with UV spectroscopy using a NanoDrop™ 2000/2000c Spectrophotometer (ThermoFisher Scientific). Wavelength measurements at 280 nm were made and protein concentration calculated using the Beer-Lambert law (Equation 3.6.1).

$$A = \varepsilon \times c \times l \tag{3.6.1}$$

where A is the light absorbance, ε is the extinction coefficient (M^{-1} cm^{-1}), c is the protein molar concentration and l is the pathlength (cm). The protein extinction

coefficients were estimated by ProtParam from the ExPASy webpage [156].

3.6.4.2 SDS-PAGE

Protein samples were mixed with SDS Loading buffer (50 mM Tris-HCl pH 6.8, 30% glycerol, 10% SDS, 2 mM DTT and 0.05% bromophenol blue) and heated to 95 °C for 5 minutes. Samples were then loaded into a 4-20% polyacrylamide gel (NuSep) in a Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (Biorad) and run at 200 V for 40 minutes in Tris-glycine buffer (25 mM Tris, 250 mM glycine and 0.1% SDS). Protein colour standard (NEB) was used as a marker to determine protein band sizes.

3.6.4.3 Western Blot

Following SDS-PAGE electrophoresis the gels was removed from the cassette and transferred to an iBlot®2 PVDF transfer stack (Invitrogen). The stack was prepared according to manufacturers instructions and ran using the preset P0 method for 7 minutes. The PVDF membrane was then transferred to 10 mL blocking buffer (3% BSA in TBST buffer (137 mM NaCl, 2.7 mM KCl and 19 mM Tris-HCl, pH 7.4)) and left to block at room temperature for 1 hour. The membrane was then left to incubate with the primary antibody diluted in blocking buffer overnight at 4 °C (Anti-CTD and Rb7 diluted to 1:1000; Anti-mouse diluted to 1:2000).

The membrane was then washed three times with TBST for 10 minutes. The membrane was then incubated with the secondary antibody at room temperature for 2 hours (anti-rabbit diluted to 1:2000). The membrane was then washed a further three times with TBST before development with either 3,3'-diaminobenzidine (DAB) or enhanced chemiluminescence (ECL).

The DAB solution was prepared by adding 25 mg of DAB to 50 mL of PBS and 33 µL of hydrogen peroxide was added immediately prior to use. 10 mL was added

to the membrane and the reaction allowed to develop for 5 minutes or until protein bands were visible. The reaction was stopped by the remove of DAB solution and resuspension of 10 mL H_2O. ECL was prepared using the Clarity^{TM} western ECL substrate kit (Biorad). 7 mL substrate mix was prepared by mixing the two substrate solutions at a 1:1 ratio. The solution was added to the blot and incubated for 5 minutes before imaging with a ChemiDoc^{TM} MP imaging system (Biorad).

3.7 NMR Spectroscopy

3.7.1 NMR Data Acquisition

NMR samples were prepared in 20 mM $NaPO_4$ buffer with 50 mM NaCl, 10% ^2H_2O, 0.05% (w/v), NaAzide and 0.1 mM PMSF. NMR experiments for RgpB-CTD resonance assignment were collected using a Bruker 700 MHz NMR spectrometer at 310 K, except for ^{15}N-HSQC-NOESY and ^{13}C-HSQC-NOESY spectra, which were collected at 800 MHz and 900 MHz, respectively. All spectrometers were equipped with a triple-resonance TXI cryoprobe.

2D HSQC experiments for ^{15}N longitudinal (T1) and transverse (T2) Relaxation Measurements were acquired as pseudo-3D experiments on a Bruker 700 MHz NMR spectrometer at 310K. Delay times of 10, 50, 100, 200, 500, 800, 1200 and 1600 ms were used to measure T1 relaxation. CPMG delay times used for T2 relaxation were 8, 16, 24, 32, 48, 64, 96, 144 ms. $\{^1$H$\}$-^{15}N NOE experiments were also measured on a Bruker 700 MHz NMR spectrometer at 310K with a 3 s saturation time.

All NMR experiments recorded for His-RgpB-CTD are listed in Table 3.7.1.

Table 3.7.1: NMR experiments recorded for RgpB-CTD.

Experiment type	Dimensions	purpose	Reference
N-HSQC	^{1}H; ^{15}N	Backbone Assignment	[171–173]
HN(CO)CACB	^{1}H; ^{13}C; ^{15}N	Backbone Assignment	[174]
HNCACB	^{1}H; ^{13}C; ^{15}N	Backbone Assignment	[145, 175]
HN(CA)CO	^{1}H; ^{13}C; ^{15}N	Backbone Assignment	[176]
HNCO	^{1}H; ^{13}C; ^{15}N	Backbone Assignment	[143, 175]
C(CCO)NH	^{1}H; ^{13}C; ^{15}N	Side-chain Assignment	[174, 177]
HC(CCO)NH	^{1}H; ^{13}C; ^{15}N	Side-chain Assignment	[174]
(H)CHH-TOCSY	^{1}H; ^{1}H; ^{13}C	Side-chain Assignment	[178]
H(C)CH-TOCSY	^{1}H; ^{13}C; ^{13}C	Side-chain Assignment	[178]
CBHD-COSY	^{1}H; ^{13}C	Side-chain Assignment	[179]
CBHE-COSY	^{1}H; ^{13}C	Side-chain Assignment	[179]
C-NOESY	^{1}H; ^{1}H; ^{13}C	Distance Restraints	[180–182]
N-NOESY	^{1}H; ^{1}H; ^{15}N	Distance Restraints	[180–182]
H-Bond HNCO	^{1}H; ^{13}C; ^{15}N	Distance Restraints	[183]
T1 Relaxation	^{1}H; ^{15}N	Dynamics	[184]
T2 Relaxation	^{1}H; ^{15}N	Dynamics	[184]
Unsat. HetNOE	^{1}H; ^{15}N	Dynamics	[184]
Sat. HetNOE	^{1}H; ^{15}N	Dynamics	[184]

3.7.2 NMR Data Processing

Raw NMR data in Bruker format was converted to NmrPipe format using NmrPipe [185]. NmrPipe functions were used to process the time-dependent data via a series of solvent filtering, zero filling, window functioning, Fourier transformation, phase correction, extraction and transposition to frequency dependent data. NMR spectra were visualised and phasing values calculated throughout this process using NmrDraw.

3.7.3 NMR Data Analysis

TopSpin 4.0.6 (academic licence) was used to process and visualise 1D NMR spectra.

CCPN Analysis 2.4 was used for visualisation, analysis and assignment of processed 2D and 3D NMR spectra [186]. The DANGLE tool in Analysis 2.4 was used to generate dihedral restraints based off backbone chemical shift assignments [149]. Hydrogen bond restraints were produced using the built-in tool in Analysis 2.4. All spectral images shown in this thesis were produced using the Analysis 2.4 print function.

T1 and T2 relaxation times were calculated using the 'follow intensity changes' tool in Analysis2.4. The fitting function $A\exp(-Bx)$ was used to calculate relaxation rates for all assigned peaks. One divided by the relaxation rate was then used to calculate the T1 and T2 relaxation time constants for each peak.

3.7.4 NMR Structural Calculation and Validation

H-bond, NOE and dihedral restraints were used by ARIA 2.3 [152], run through the NMRBox server [187], to create structural model ensembles. Eight iterations, with 20 conformers per iteration, were performed. Water refinement was the performed for the 10 lowest energy state models from the eighth iteration ensemble. A total of 2207 unambiguous NOE distance restraints, 524 ambiguous NOE distance restraints, 22 hydrogen bond restraints and 92 dihedral restraints were used for the ARIA structural

modelling.

Structure quality was calculated using ARIA reports and Procheck-NMR [188]. The online server RAMPAGE was used to produce Ramachandran plots for the water refined structural ensemble[189]. Structure visualisation was performed using PyMOL [190]. All structural images were produced in PyMOL using ray-tracing drawing.

3.8 Protein-Protein Interaction Studies

3.8.1 *P. gingivalis* Growth Conditions

Both plate and liquid cultures of *P. gingivalis* were grown in a Don Whitley anaerobic workstation to maintain anaerobic conditions and a constant temperature of 37 °C. Blood agar plates were grown for 5 days or until colonies could be visibly identified. For recently electroporated *P. gingivalis* cells, it would often take 10-14 days before colonies would emerge. BHI-H cultures were grown for 24 hours under these conditions with no shaking agitation.

3.8.2 Fractionation of Bacterial Cultures

P. gingivalis cells were collected by centrifugation at $10,000 \times \mathbf{g}$ for 10 minutes at 4 °C and the pellet resuspended in PBS, supplemented with 5 mM Nα-Tosyl-L-lysine chloromethyl ketone hydrochloride (TLCK) and 10 mg mL^{-1} Leupeptin. The supernatant was retained as the culture fluid fraction. Cells were lysed by sonication for 5 minutes at 70% power with 3 second pulsing. The lysate was centrifuged again at $10,000 \times \mathbf{g}$ for 10 minutes to collect cellular debris.

Extraction of periplasmic proteins from *P. gingivalis* was performed using osmotic shock [108]. 20 mL of *P. gingivalis* culture, grown for 16 hours to an OD_{600}~ 1.5, was centrifuged at $10,000 \times \mathbf{g}$ for 30 minutes at 4 °C. A sample of the culture fluid was

collected for analysis by SDS-PAGE. The pellet was washed once with 5 mL PBS to remove excess culture media and cells pelleted again. The pellet was then suspended in 500 mM sucrose, 20 mM Tris-HCl, pH 7.4, 5mM TLCK and 1× cOmplete™ EDTA-free protease inhibitor cocktail (PIC; Sigma Aldrich). The solution was incubated for 10 minutes on ice followed by centrifugation at $8000 \times \mathbf{g}$ for 30 minutes at 4 °C. The cell pellet was immediately resuspended in 5 mL ice cold buffer containing 5 mM $MgCl_2$, 1 mM TLCK and 1× PIC and incubated for 10 minutes on ice. The suspension was centrifuged at $10,000 \times \mathbf{g}$ for 30 minutes at 4 °C and the supernatant taken as the periplasmic fraction. Cell pellet was taken as the inner-membrane and cytoplasmic fraction.

3.8.3 Ni-NTA Pull Down

Pull-down experiments were performed using recombinantly expressed His-tagged RgpB-CTD, purified by Ni-nitrilotriacetic acid (NTA) chromatography and Size-exclusion chromatography.

For reactions in which no cross-linking was performed, 50 µm His-RgpB-CTD was first captured using 40 µl pre-equilibrated Ni-NTA agarose resin. This was washed three times with using 200 µl Ni-NTA wash buffer supplemented with triton-X100, to remove unbound protein. Cells from a 10 mL *P. gingivalis* collected and lysed by sonication, as described in Section 3.8.2. The buffer used for sample lysis was supplemented with 1% triton X-100 to solubilise membrane proteins. Lysates were incubated with the Ni-NTA:His-RgpB-CTD suspension for 1 hour at 4 °C with gentle mixing.

For samples where chemical cross-linking was incorporated into the protocol, His-RgpB-CTD was added directly to *P. gingivalis* lysate and formaldehyde added to a final concentration of 1%. The suspension was incubated for 30 minutes at 37 °C

followed by the addition of 1M glycine to a final concentration of 125 mM to quench the cross-linking reaction. Quenching was allowed to proceed for 15 minutes at room-temperature and then urea added to a final concentration of 8M and SDS to a final concentration of 1% to denature all proteins. The sample was loaded onto 40 µl pre-equilibrated Ni-NTA agarose resin.

The sample was washed for 15 minutes at 4 °C three times, using 5 column volumes of Ni-NTA wash buffer. wash buffer supplemented with 8M ureas and 1% SDS was used for pull-down experiments containing the additional cross-linking step. Samples were eluted directly into 80 µl SDS-PAGE loading buffer and heated to 70 °C for 20 minutes.

3.8.4 pTIO-AI Protein Extraction

Extraction of protein from *P. gingivalis* strains carrying pTIO-AI-ss-CTD-His, pTIO-AI-ss-His-CTD or pTIO-AI-CTD-His was performed using whole cell lysates, periplasmic fractions and clarified culture fluid. The method of fractionation is described in Section 3.8.2. Samples from each fraction were purified using Ni-NTA agarose resin (ThermoFisher Scientific) with a batch flow protocol. 50 µl of resin was loaded into a 1.5 mL microcentrifuge tube and equilibrated using 10 resin bed volumes of Ni-NTA wash buffer. 1 mL of fractionation sample was loaded onto the resin and incubated at 4 °C for 1 hour with gentle mixing. The resin was then washed three times with 250 µl Ni-NTA wash buffer. At each washing step the sample was incubated at 4 °C with gentle mixing for 15 minutes. Elution was achieved by adding 50 µl 2×SDS-PAGE loading buffer directly to the resin followed by heating to 95 °C for 5 minutes. Samples of the Ni-NTA chromatography flow-through, wash and elution were also taken for SDS-PAGE analysis.

3.8.5 Co-Immunoprecipitation

Co-immunoprecipitation reactions were performed using the BioVision Immunoprecipitation Kit according to manufacturer's instructions. 20 mL *P. gingivalis* cultures were centrifuged at $10,000 \times \mathbf{g}$ for 10 minutes at 4 °C to harvest bacterial cells. Pellets were washed three times using 5 mL PBS and finally resuspended in 1 mL PBS. Where appropriate, 50 mM DSP was added at a final concentration of 2 mM, or formaldehyde added to a final concentration of 1%. Cross-linking reactions were incubated at 37 °C for 30 minutes followed by 15 minute quenching via the addition of 100 mM Tris, pH 7.4.

For reactions in which no chemical cross-linking was performed the non-denaturing lysis buffer, supplemented with 5 mM TLCK, was used for protein extraction. For reactions involving chemical cross-linking the denaturing RIPA lysis buffer was used. 50 µl of 1 mg mL^{-1}Anti-RgpB-CTD Rabbit immunoglobulin (Generon) was added to each sample and left to incubate overnight at 4 °C with gentle mixing.

30 µl of pre-equilibrated Protein A/G beads were used per sample for antibody capture and suspensions incubated at 4 °C for 1 hour. Beads were washed three times using 1 mL of kit wash buffer and all samples were eluted using a denaturing method via the addition of 40 µl 2×SDS-PAGE loading buffer.

3.9 Molecular Dynamics and Protein Docking

3.9.1 Production of Protein Models

The cryo-EM model for *F. johnsonsiae* PorV:SprA complex (PDB: 6H3I [122]) was initially ran through the online MODELLER server [191] to fill missing atoms in the structural model. The Phenix package, Cryo_fit1, was used to further refine the structural model via flexible fitting of the structural model to the electron microscopy

map by molecular dynamics [192]. The default values were used for all fitting options. The resulting structural model of PorV was selected and used for molecular dynamics simulations.

Given that there is no experimental structural model of *P. gingivalis* PorV, a predicted structural model was derived using the following method. Firstly, the amino acid sequence for *P. gingivalis* W50 PorV (excluding the N-terminal Sec signalling peptide) was uploaded to the online server GREMLIN [193] for the generation of co-evolutionary distance restraints. The GREMLIN multiple sequence alignment was generated using the HHBlits method and an E value cutoff of 1e-10 selected. The resulting distance restraints were used alongside the primary protein sequence to perform comparative modelling using the Robetta protein structure prediction server [194]. A total of 100 structural models was generated and the *F. johnsonsiae* PorV structural model used as a template for modelling.

3.9.2 Molecular Dynamics Simulations

The *F. johnsoniae* experimental PorV model and the *P. gingivalis* predicted PorV model were first run through the online PPM server [195] to determine the rotation and position within the membrane. The output from this server was then uploaded the CHARMM-GUI membrane builder server [196] to generate the necessary input files to build the protein:membrane complex for molecular dynamics simulations. A heterogeneous lipid bilayer with 100 lipids per leaflet was created using 80% 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) and 20% 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE). The initial system size was $90 \times 90 \times 130$Å for both models and the system was determined using the replacement method with 150 mM NaCl.

Simulations were performed using GROMACS with the CHARMM36m force field.

The system was minimised and equilibrated at a temperature of 310K and a pressure of 1 bar using protocol files produced using CHARMM-GUI [197]. Once equilibrated a 200 ns production simulation was carried out at 310K and a pressure of 1 bar. The temperature was controlled using the Nose-Hoover thermostat and the pressure was controlled using the Parrinello-Rahman barostat.

The trajectory output was centred and fitted around the protein using the 'gmx trjconv' command. The python package MD-Analysis [198] was used for analysis of trajectory RMSD and RMSF. Clustering of frame states was performed using the python package TTClust used for the clustering a trajectory frame states [199]. The clustering analysis was performed using Cα atom RMSD values and an empirically determined cutoff value was used to generate representative clusters. Representative models from each cluster were combined into a structural ensemble for protein docking experiments. Secondary structure analysis was performed using the VMD Timeline module [200].

3.9.3 Protein Interaction Docking

For protein interaction docking of *P. gingivalis* PorV and RgpB-CTD the cluster ensemble from TTClust was used. For RgpB-CTD the ten water refined NMR structural models were aligned using POSA server [201] and the RMSD values for each model was summed. The model with the lowest sum-RMSD value was taken as representative model for the structural ensemble. Representative *P. gingivalis* PorV and RgpB-CTD structural models were uploaded to the CPORT server [202] for the prediction of protein-protein interface residues. For PorV, predicted residues that are membrane embedded or at the inner leaflet were excluded.

The PorV structural ensemble and the RgpB-CTD structure were submitted to the HADDOCK2.4 server [203] for flexible docking analysis. The list of refined CPORT

residues was used to assign active residues. Passive residues were defined as those within 6.5Åof active residues. The structures sampled in iteration 0, 1 and water-refinement was set to 10,000, 400 and 400, respectively. A total of five trials was used for rigid body minimisation with 180° rotational sampling enabled.

3.10 Statistical Analysis

The statistical analysis for detergent extraction screening was adapted from Morrsion *et al.,* [169] and was carried out in Python 3.6 using the Statsmodels v0.12.1 module [204] and Scipy library [205]. The data was fitted using the ordinary least squares method, with solubilisation agent as the independent variable. A 1-way ANOVA was then performed on the fitted data and the difference between individual groups was analysed by a Tukey *post hoc* test. A value of $P < 0.05$ was considered to be statistically significant. A Levene test was used to calculate the equality of variance between groups and a Shapiro-Wilk test was used to calculate normality for each group.

Chapter 4

Structural Characterisation of the RgpB C-terminal Domain

4.1 Chapter Aims

The aim of this chapter was to understand the structural nature of the CTD of cargo proteins secreted by the T9SS. To date, there have been 35 proteins identified in *P. gingivalis* which possess a CTD recognised by the T9SS [81]. When this project was started there were no structural models available for a T9SS cargo CTD. In this chapter, the structure of one CTD from a T9SS cargo protein was characterised in detail using NMR Spectroscopy. This model was produced to further our understanding of how this domain can facilitate cargo protein export and the identification of structural features that may allow this.

NMR spectroscopy was selected as it provides a model for the protein in solution state, which more closely resembles physiological conditions compared to X-ray crystallography. In addition, NMR spectroscopy allows experimental studies to be performed for protein dynamics and for detection of protein-protein interactions. Given that this project ultimately seeks to understand how the CTD interacts with compon-

ents of the T9SS this makes the technique particularly well suited to the aims set out here.

4.2 Rational Design of Recombinant RgpB C-Terminal Domain

4.2.1 Selection of Model CTD

There are a number of published studies that investigate the consensus CTD sequence recognised by the T9SS [80, 81, 97, 206, 207]. Veith *et al.,* [81] used 30 experimentally identified CTD proteins from *P. gingivalis* and *Tannerella forsythia* for analysis by the GLAM2 [208] and MEME [209] sequence alignment servers to identify conserved motifs in the CTD of T9SS cargo proteins. The author's analysis revealed a conserved motif of 42 amino acids containing three conserved motifs (B, D and E) identified by Seers *et al.*, [80]. Motifs B and D could readily be identified using multiple sequence alignment for all known CTD proteins from the *P. gingivalis* W50 strain using the online MUSCLE server [163] (Figure 4.2.1A). However, conservation in motif E, originally reported to be seven residues in length, was more difficult to identify.

It has previously been demonstrated that the last 22-residues of the Hbp35 CTD, containing motifs D and E, was sufficient to achieve secretion, supporting the theory that these motifs are important in recognition by the secretion system [97]. However, it has since been suggested that structural conservation is more important than sequence conservation in allowing cargo protein recognition [98]. Therefore, determining a structural model for the domain would prove useful to functional studies and modelling of interaction complexes. In this study, the CTDs from the three gingipains, RgpA, RgpB and Kgp were selected because these are the most widely studied T9SS cargo proteins with crystal structures available for the catalytic domains of RgpB [210]

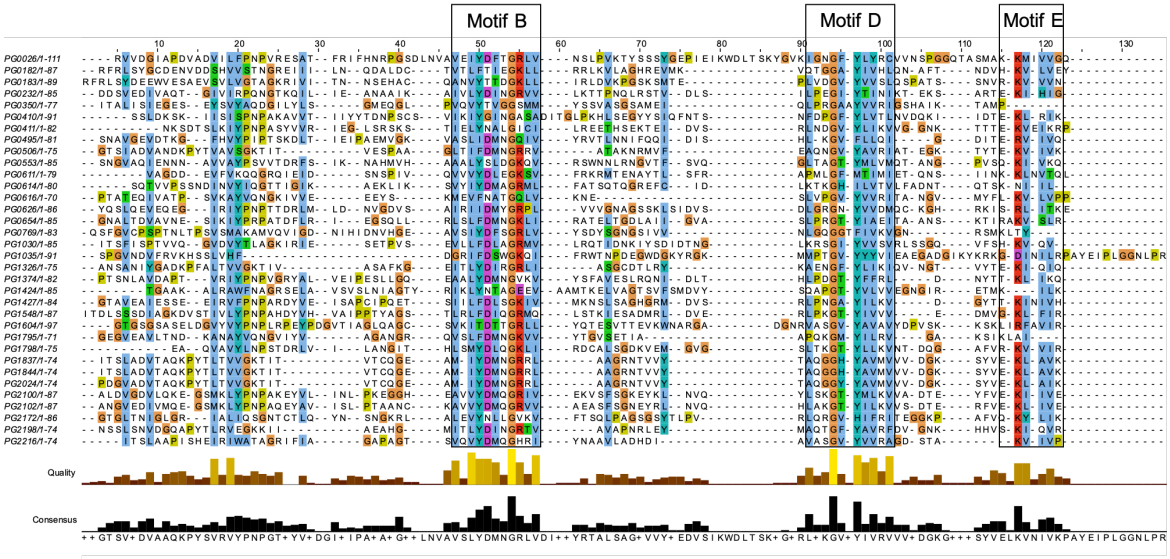


Figure 4.2.1: Multiple sequence alignment of *P. gingivalis* W50 gingipain C-terminal domains using MUSCLE. The three conserved sequence motifs (B, D and E) identified by [81] are highlighted.

(PDB: 4IEF) and Kgp [52] (PDB: 4RBM). In addition, it has been shown that the gingipains are the most abundant T9SS cargo proteins in the outer membrane and outer membrane vesicles of *P. gingivalis* [44, 84].

Alignment of the three gingipain CTDs shows that Kgp-CTD and RgpA-CTD have 95.8% sequence identity. The RgpB-CTD has 50.0% and 52.7% identity to RgpA-CTD and Kgp-CTD, respectively. All three gingipains contain the three conserved motifs B, D and E. Given the sequence identity between these three proteins, RgpB was selected for further analysis and recombinant expression as it is perhaps the most extensively studied gingipain.

An NCBI BlastP [162] search using the *P. gingivalis* W50 RgpB-CTD sequence, with an expect threshold of 0.05, was used to identify 322 aligned sequences. The Clustal Omega server [164] was then used to generate a multiple sequence alignment for the BlastP results and a consensus RgpB-CTD sequence was generated using the online WebLogo server [165] (Figure 4.2.2).

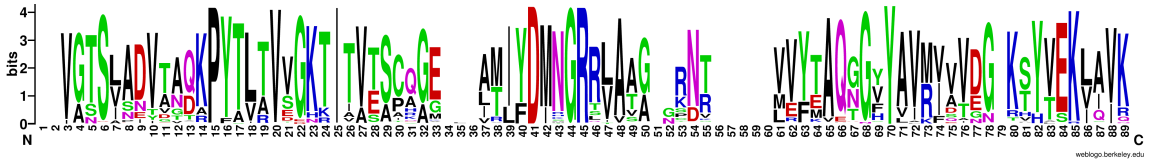


Figure 4.2.2: Consensus CTD sequence for *P. gingivalis* W50 RgpB. The Alignment was generated using NCBI BlastP [162] and Clustal Omega [164]. The consensus sequence image was generated using the WebLogo server [165].

4.2.2 Construct Design

The *P. gingivalis* strain W50 (ATCC 53978) was selected as a model organism for this study. W50 was selected as it is a virulent and prevalent strain, making it a clinically relevant strain to study [211]. In addition, a variety of deletion mutants have been created using this strain which are available in our laboratory. Access to this wider array of mutant strains has provided more options for phenotypic screening throughout the research project.

The crystallographic structure of mature RgpB, purified from *P. gingivalis,* has S^{664} as the C-terminal residue, suggesting this is the site of CTD cleavage [79] (PDB: 1CVR). However, mass spectrometry analysis suggests the CTD cleavage site is at T^{663} [82]. To ensure the cleavage site is included in recombinant constructs the sequence G^{662}- K^{736} was taken from the *P. gingivalis* W50 rgpB gene (PG0506; NCBI Reference Sequence: WP_010956050.1) and cloned into the recombinant expression vector pET46 EK/LIC using ligation independent cloning (Figure 4.2.3). This vector adds a 6×His-tag to the N-terminus of the protein, followed by an enterokinase cleavage site to allow tag removal. The vector possesses a T7 promoter with *lac* operator control, allowing IPTG induced expression, and an Amp^R cassette for positive selection. These features make this vector an ideal choice for recombinant expression in *E. coli.*

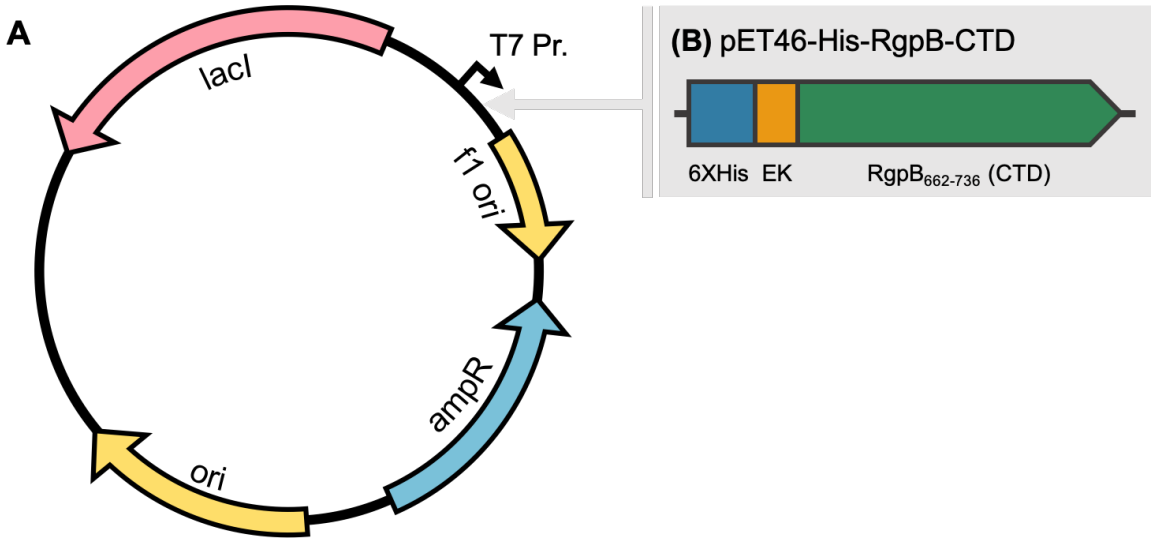


Figure 4.2.3: Vector map of pET46-RgpB-CTD. (A) Circular plasmid map for pET46 EK/LIC vector. (B) Open reading frame of RpgB-CTD (G^{662}-K^{736}) insert with N-terminal His-tag and enterokinase cleavage site under the control of the IPTG induced T7 promoter.

4.3 RgpB-CTD Expression and Purification

4.3.1 Protein Expression

The *E. coli* strain BL21 (DE3) was selected for expression due to its compatibility with the T7 expression system, high yields and ease of use [212]. Cells were grown to an OD_{600} of 0.8 in LB media at 37 °C with 240 rpm shaking. Cells were induced with 100 µM IPTG and left for 16 hours at 18 °C with 240 rpm shaking. Preparations using these conditions consistently yielded around 20 mg L^{-1} of culture. Therefore, no further optimisation of these expression conditions was performed.

4.3.2 Ni-NTA Chromatography

Cells were collected by centrifugation, resuspended in lysis buffer the sample and lysed by sonication (Section 3.6.1.2). Following sonication, the cytosolic fraction was obtained by centrifugation, and loaded onto a 5 mL HisTrap column. After washing, the

protein was eluted using a linear gradient of low to high imidazole, which competitively binds the NTA resin, releasing the His-tagged protein. Fractions containing His-RgpB-CTD were confirmed through SDS-PAGE analysis, and these fractions pooled. For long-term storage, some preparations were lyophilised following His-Trap purification. Prior to this, a PD-10 buffer-exchange column was used to remove the imidazole from the sample. Figure 4.3.1 shows this protocol was successful in the purification of His-RgpB-CTD with minimal contamination and degradation.

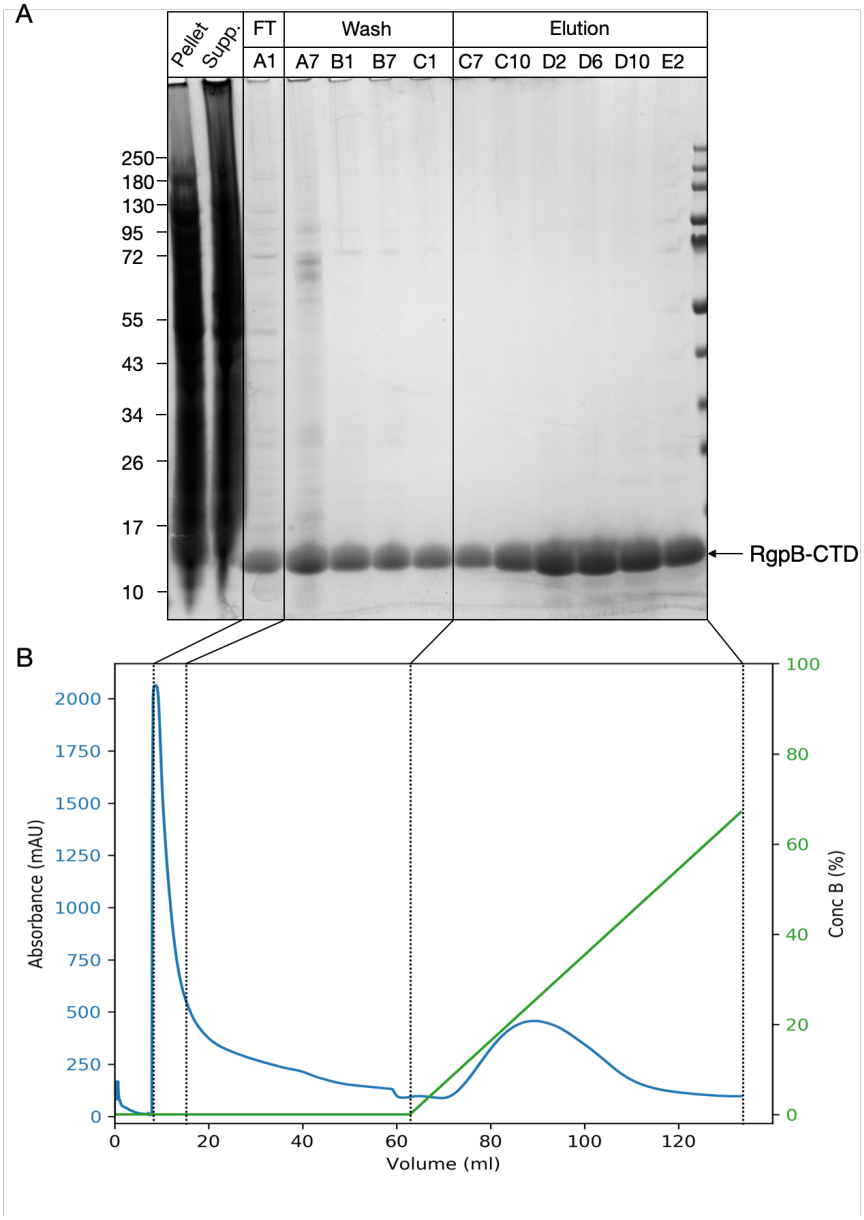


Figure 4.3.1: HisTrap purification of His-RgpB-CTD. (A) SDS-PAGE of HisTrap elution profile. (B) UV A280 absorbance chromatogram (blue) with elution gradient from buffer A (20 mM imidazole) to buffer B (300 mM imidazole) (green). Fractions C7-E2 were deemed pure enough for pooling and further study. Legend: Pellet - Insoluble cell lysate pellet, Supp. - Soluble cell lysate fraction, FT - Flow-through during sample loading, A1-E2 - 2 mL fractions collected by AKTAPure.

4.3.3 Size Exclusion Chromatography

Pooled sample from Ni-NTA purification was further purified by size exclusion chromatography to remove any additional contaminants and to transfer the protein to a suitable buffer for downstream assays. Due to the molecular size of the His-RgpB-CTD construct (9.8 kDa), a HiLoad 16/600 Superdex 75 pg column was used as it provides good resolution for proteins between 10-75 kDa. Figure 4.3.2A shows that His-RgpB-CTD elutes as a single major peak. Separation of collected fractions comprising this peak shows it is a single clean band of the correct size with only minimal degradation (Figure 4.3.2B).

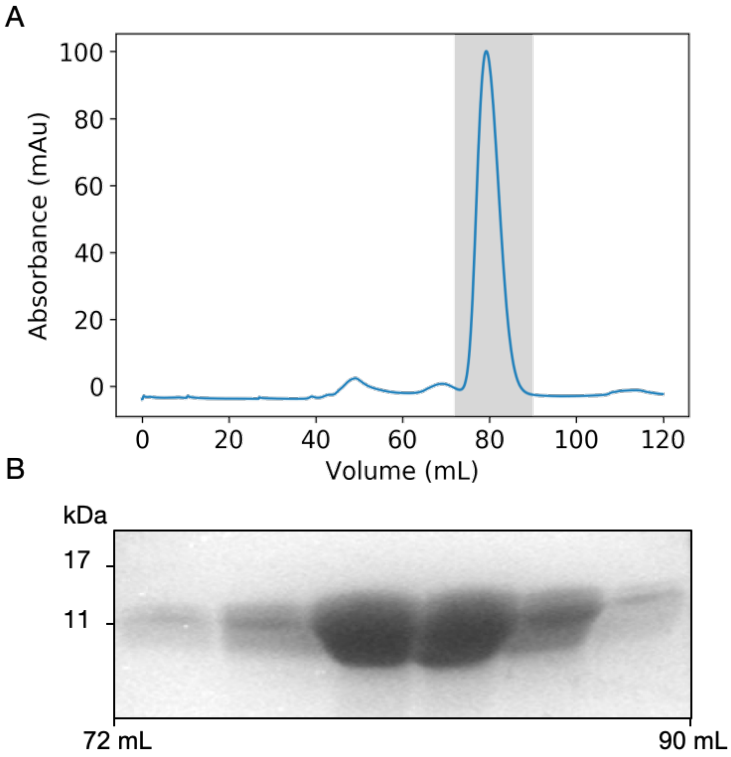


Figure 4.3.2: Result of size exclusion chromatography for His-RgpB-CTD. (A) UV A280 absorbance chromatogram with the His-RgpB-CTD peak highlighted in grey. (B) SDS-PAGE of the His-RpgB-CTD 3 mL fractions from 72 mL to 90 mL highlighted in grey on the above chromatogram.

4.4 Initial NMR Characterisation of RgpB-CTD

Following purification, 1D ^{1}H NMR was used for the initial characterisation of His-RgpB-CTD to determine its suitability for NMR structural studies. A spectrum for 1D His-RgpB-CTD, recorded at 310 K, is shown in Figure 4.4.1. The spectrum displays a good dispersion of chemical shifts in the region of 10-6 ppm, characteristic of amides forming ordered secondary structure elements in the protein. In addition, peaks present between 2-0 ppm indicates the methyl and aromatic resonances are forming a structured protein core.

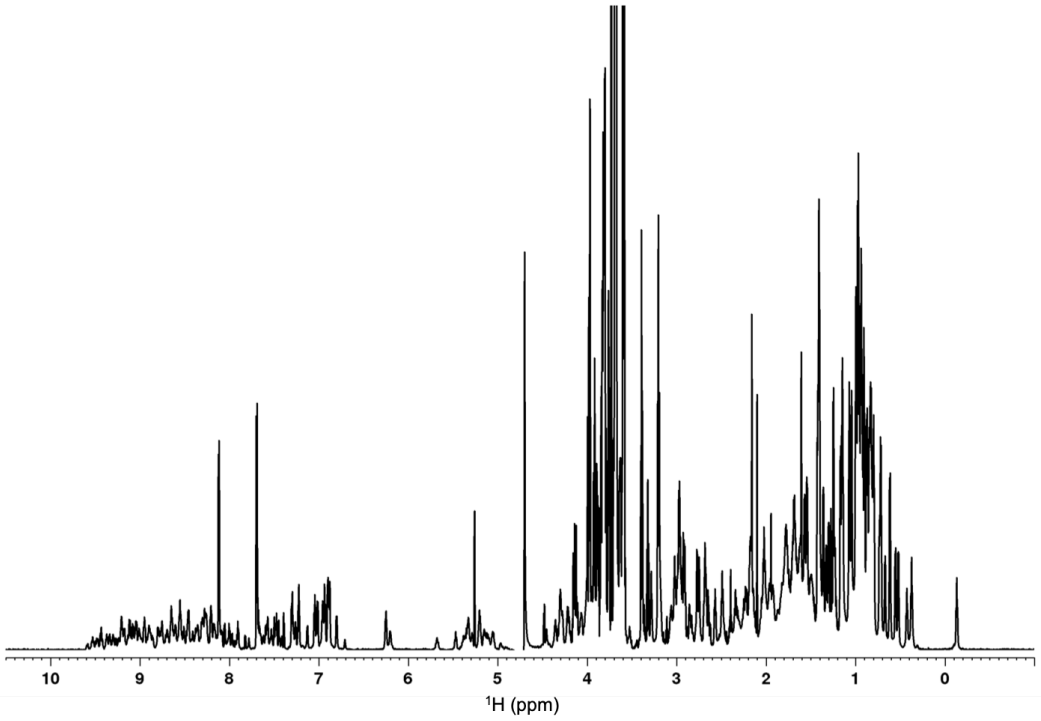


Figure 4.4.1: ^{1}H NMR spectrum of RgpB-CTD. The Spectrum shows good dispersion of amide chemicals shifts between 10-6 ppm and methyl region peaks between 2-0 ppm, indicative of a well-folded protein. The spectrum was recorded using a Bruker 700 MHz NMR spectrometer at 310 K.

To further determine if the His-RgpB-CTD construct would remain stable for the duration of the runtime of longer NMR experiments, required for full assignment and NOE restraints acquisition, an ^{15}N labelled sample was prepared for 2D ^{1}H-^{15}N

heteronuclear single quantum coherence (HSQC) spectroscopy [171–173, 213]. This experiment was measured at 310 K (37 °C) because this temperature is physiologically relevant for *P. gingivalis* pathology and higher temperatures cause the protein to tumble faster in solution giving narrower line-widths. The ^{15}N-HSQC spectrum was recorded for His-RgpB-CTD after 0, 2.5 and 5 days incubation at 37 °C. As shown in Figure 4.4.2 there is no observable change in the spectrum over the five-day time-course. In addition, peaks are well dispersed with narrow line widths indicative of a well-folded protein. There were only a few overlapping peaks, primarily in the centre of the spectrum, which were later found to mostly correspond to the highly disordered N-terminal His-tag.

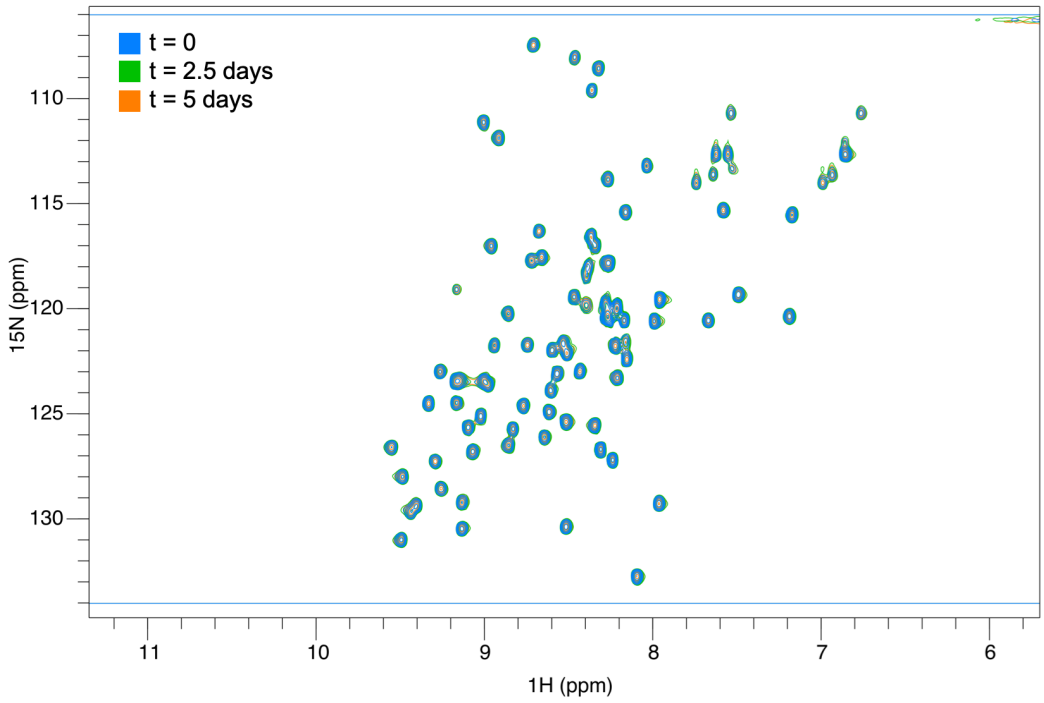


Figure 4.4.2: 2D ^{1}H-^{15}N-HSQC analysis of His-RgpB-CTD temperature stability. Spectra were recorded for the protein after 0, 2.5 and 5 days incubation at 37 °C (310K). Good dispersion and narrow line widths could be observed for the majority of peaks suggesting this is a well-folded protein. No new peaks were observed following 5 days incubation at 37 °C, showing that the protein does not significantly degrade during this time. The spectra were recorded using a Bruker 700 MHz NMR spectrometer at 310 K.

4.5 RgpB-CTD Chemical Shift Assignment

4.5.1 Backbone Resonance Assignment

The assignment of backbone resonance chemical shifts is a critical step in studying proteins by NMR as it allows for the sequential identification of residues in a protein. It is also necessary for the analysis of more advanced spectra that show side-chain resonances or for the generation of distance restraints. To begin the assignment process 79 peaks were selected, excluding side-chain peaks, from the ^{15}N-HSQC spectrum shown in Figure 4.4.2. A total number of 87 peaks were expected, however, due to overlapping peaks for the N-terminal His-tag and enterokinase site in the centre of the spectrum not all peaks could be resolved.

Triple resonance experiments (HN(CO)CACB, HNCACB, HNCO and HN(CA)CO [214]) were recorded to allow sequential backbone assignment for ^{13}C ^{15}N labelled His-RgpB-CTD. The ^{15}N and ^{1}H chemical shift assignment of the ^{15}N-HSQC peaks were used to guide peak identification in the triple resonance spectra. HNCACB and HNCO experiments show intra- and inter-residue chemical shifts, whereas the HN(CA)CO and HN(CO)CACB experiments only provide inter-residue chemical shifts. By matching the corresponding peaks between the HN(CA)CO to the HNCO and the HN(CO)CACB to the HNCACB it is possible to assign C_α, C_β and C' chemical shifts for each ^{15}N ^{1}H spin system.

The selected peaks from the ^{15}N HSQC spectrum were assigned to individual spin systems with an arbitrary numbering system. The semi-automated 'protein sequence assignment' tool from Analysis 2.4 [186] was then used to align spectral peaks in the triple resonance spectra. The program uses the assigned spin systems to identify peak positions in the triple resonance spectra. These peaks in the NH(CO)CACB and HN(CA)CO are then used to query the HNCACB and HNCO, respectively. By

aligning the inter-residue peaks of the HN(CO)CACB with intra-residue peaks of the HNCACB strip of a different spin system to determine a connection between the two spin systems. The HN(CA)CO-HNCO assignment is performed in tandem to help elucidate more ambiguous connections. An example of the sequential assignment process, using the NH(CO)CACB and HNCACB spectra, from residues V702-A705 is shown in Figure 4.5.1A. The corresponding assignment of V702-A705 C' resonances, using the HNCO and HN(CA)CO spectra, is shown in Figure 4.5.1B.

Using sequential assignment it was possible to achieve 100% assignment of backbone resonances ($^{1}H_{N}$, ^{15}N, C_{α}, C_{β} and C'), excluding the His-tag, enterokinase cleavage site and proline resonances. Full assignment of prolines is not possible due to the cyclic secondary amide structure of proline residues.

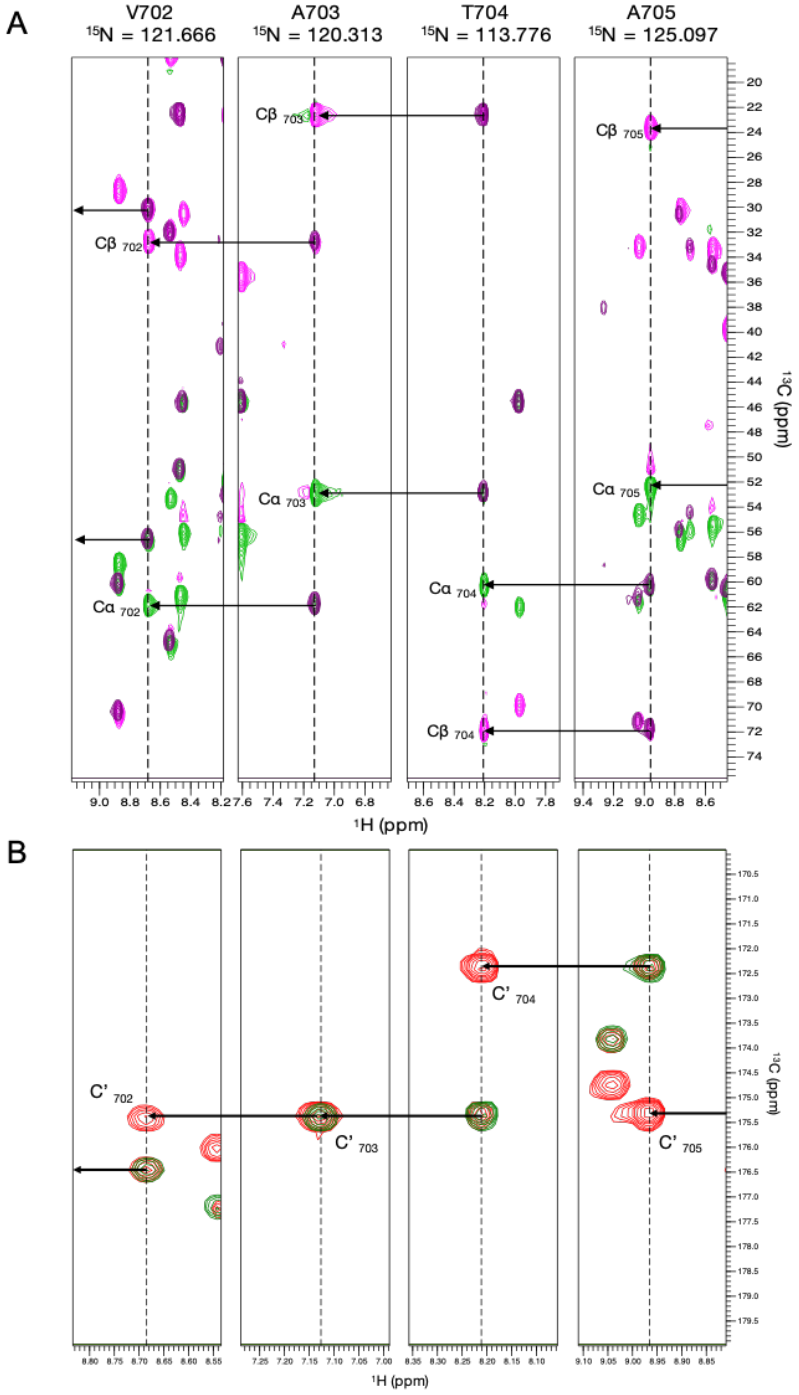


Figure 4.5.1: Sequential assignment example for residues V702 to A705 in His-RgpB-CTD. (A) C_α and $C_\beta$assignment using NH(CO)CACB(purple) and HNCACB (green, positive/pink, negative). (B) C' assignment using HNCO (red) and HN(CA)CO (green) spectra. Sequential matches are denoted with a black arrow. The spectra were recorded using a Bruker 700 MHz NMR spectrometer at 310 K.

All resonances comprising a given spin system have distinct chemical shift values that are dependent on amino acid type. It is therefore possible, following chemical shift assignment, to determine amino acid types for spin systems with backbone resonance assignments. Using reference chemical shift tables in Analysis 2.4, the arbitrarily numbered spin systems were assigned to specific amino acid types. These assignments were used alongside the sequential links derived from the triple-resonance spectra, to calculate residue positions within the His-RgpB-CTD protein. A fully assigned ^{15}N HSQC spectrum is shown in Figure 4.5.2.

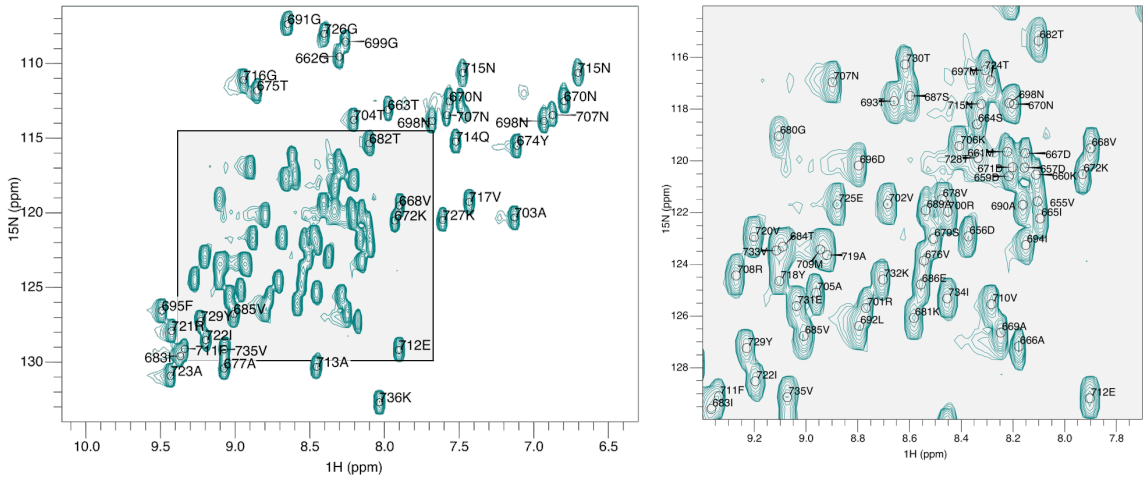


Figure 4.5.2: Assigned 2D ^{1}H-^{15}N-HSQC spectrum of His-RgpB-CTD. Assigned ^{1}H-^{15}N peaks are annotated with residue number and amino acid type. The spectrum was recorded using a Bruker 700 MHz NMR spectrometer at 310 K.

4.5.2 Side-chain Resonance Assignment

Side-chain assignments were initially mapped using spectra from the C(CCO)NH and HC(CCO)NH experiments. These spectra show resonances for aliphatic side-chain carbons and hydrogens, respectively, of the preceding residue. The ^{15}N and ^{1}H chemical shift assignments from the ^{15}N-HSQC are again used to guide assignment. As shown in Figure 4.5.3 a ^{15}N plane is navigated to and the ^{1}H assignment used to guide

assignment of $C_{\alpha/\beta/\gamma/\delta/\epsilon}$ and $H_{\alpha/\beta/\gamma/\delta/\epsilon}$ resonances. Cα and Cβ chemical shifts derived from the NH(CO)CACB and HNCACB spectra were also used to assist assignment.

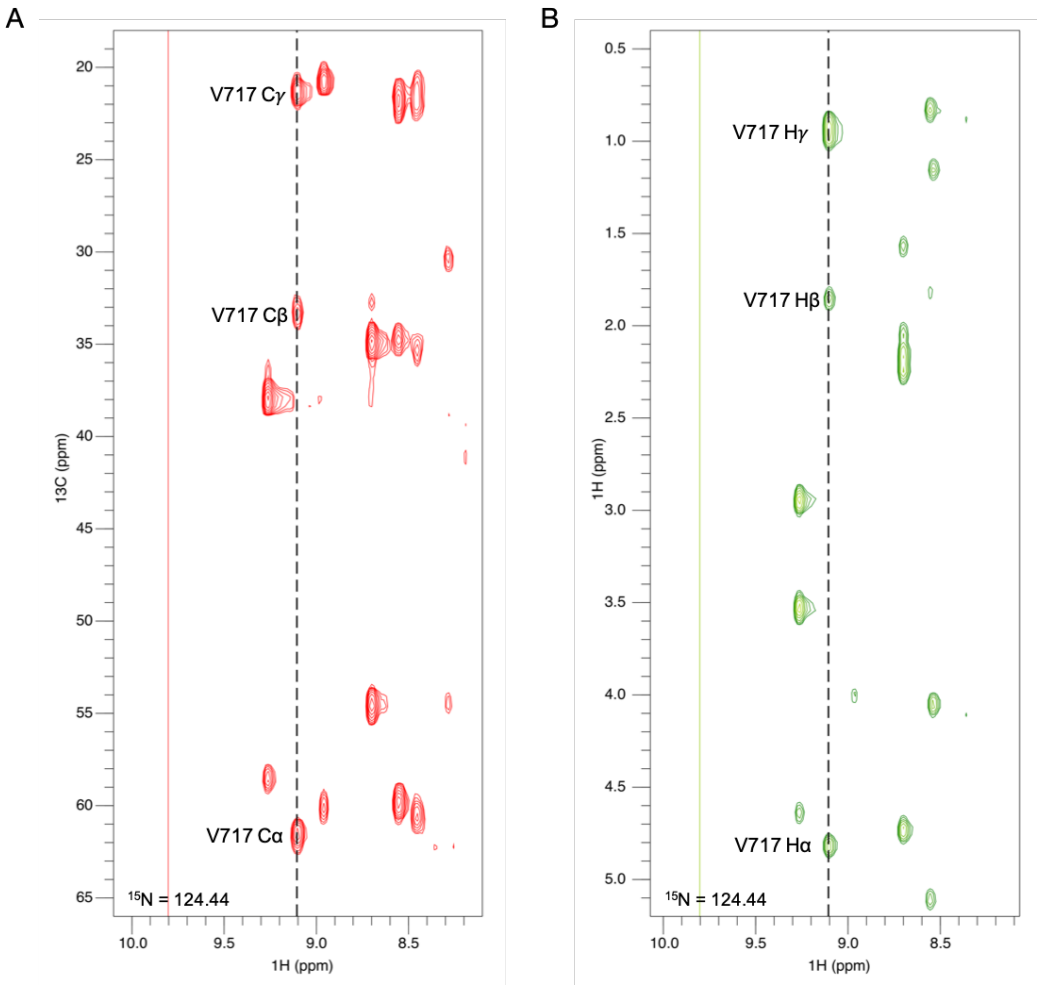


Figure 4.5.3: Example strips for side-chain ^{13}C and ^{1}H assignment using HC(CCO)NH C(CCO)NH spectra. The N and H planes from Y718 are shown. Peaks represent $C_{\alpha/\beta/\gamma/\delta/\epsilon}$ and $H_{\alpha/\beta/\gamma/\delta/\epsilon}$ resonances from the preceding V717 residue. (A) C(CCO)NH with Cα , Cβ and Cγ peaks of V717 highlighted. (B) HC(CCO)NH with Hα , Hβ and H$\gamma$of V717 highlighted. The spectra were recorded using a Bruker 700 MHz NMR spectrometer at 310 K.

Assignment of the C(CCO)NH and HC(CCO)NH is not strictly necessary for the full assignment of side-chain resonances as this can be solely achieved using the (H)CHH-TOCSY (^{13}C resonances) and the H(C)CH-TOCSY (^{1}H resonances) spectra. However, the C(CCO)NH and HC(CCO)NH assignments help to initiate the

(H)CHH/H(C)CH-TOCSY assignment process by providing a greater starting number of side-chain resonance assignments. This is useful when trying to resolve ambiguities and assignment of peaks that are weak or overlapping. The H(C)CH-TOCSY spectra provide strips for each ^{13}C and ^{1}H resonance in the protein side-chain with all side-chain hydrogen resonances visible. The (H)CCH-TOCSY provides similar information but instead shows all side-chain carbon resonances. An example strip for the Cδ - Hδ resonances of R701 is shown in Figure 4.5.4.

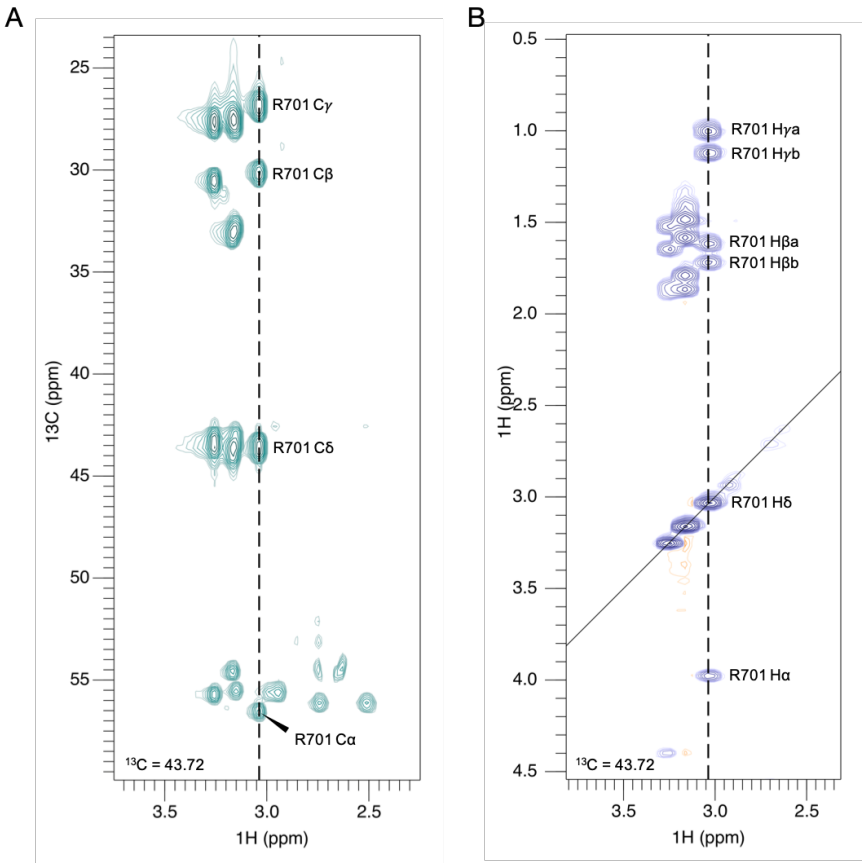


Figure 4.5.4: Assignment of example strips from (H)CHH-TOCSY and H(C)CH-TOCSY spectra. (A) A single ^{1}H-^{13}C strip in the (H)CHH-TOCSY showing the assignment of side-chain ^{13}C resonances for residue R^{701}. (B) ^{1}H-^{13}C strip in the H(C)CH-TOCSY showing side-chain ^{1}H resonance assignment for R^{701}. The spectra were recorded using a Bruker 700 MHz NMR spectrometer at 310 K.

CBHD and CBHE experiments were also recorded for His-RgpB-CTD to allow

assignment of aromatic Hδ and Hϵ resonances of tyrosine and phenylalanine residues. Given that these aromatic rings often form the core of folded protein assignment of these resonances is beneficial to structural calculation. This is because resonances in the core of the protein show a high number of distance restraints which helps to guide the calculation of structural models. The CBHD and CBHE spectra were assigned using the known Cβ shifts and reference chemical shift charts. The fully assigned spectra can be seen in Figure 4.5.5. It was possible to assign all tyrosine and phenylalanine Hδ and Hϵ resonances using these spectra.

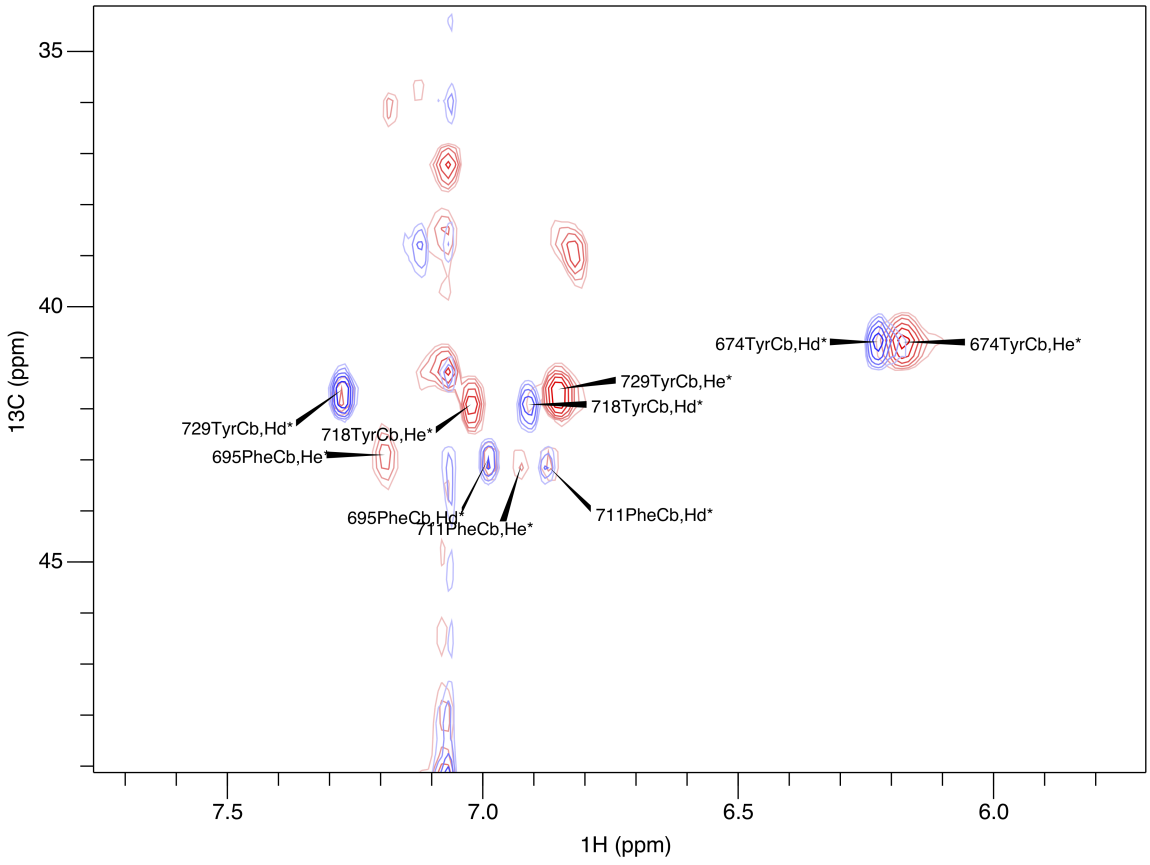


Figure 4.5.5: Assignment of His-RgpB-CTD HD and HE resonances using CBHD and CBHE spectra. The CBHD spectrum, used for assignment of HD resonances, is shown in blue. The CDHE spectrum is shown in red and was used for assignment of HE resonances. The spectra were recorded using a Bruker 700 MHz NMR spectrometer at 310 K.

Reference chemical shift tables in Analysis 2.4 were used to assign atom type to peaks in the assigned C(CCO)NH, HC(CCO)NH, H(C)CH-TOCSY and (H)CCH-TOCSY spectra. Using this method it was possible to assign 253 side-chain ^{1}H resonances (96.2%) and 165 side-chain ^{13}C resonances (85.1%). The reason for incomplete assignment, particularly of ^{13}C resonances is because it is not possible to measure the resonances of certain side-chain atoms using the spectra mentioned here. For example, the Cγ resonance of aspartate is not bonded to any ^{1}H resonances and therefore cannot be detected in any of the above experiments.

4.6 RgpB-CTD Structural Restraints

4.6.1 Dihedral Angle Restraints

The estimation of phi and psi dihedral angles allow the prediction of protein secondary structure. These restraints are also used during automated structure calculation, alongside distance restraints, to guide the algorithm. The DANGLE (dihedral angles from global likelihood estimates) tool [149] within Analysis 2.4 was used for estimation of dihedral angle restraints. The program utilises amino acid sequence, chemical shift assignments and a database of chemical shifts for known protein structures to predict phi and psi angles. A maximum of two discontinuous prediction islands was selected to filter out poorly predicted restraints.

A total of 92 restraints were generated, composed of 46 phi and 46 psi restraints. Dihedral restraints for the remaining 41 residues were not committed as they exceeded the maximum number of prediction islands. All rejected restraints were in regions of the protein that were predicted to be disordered.

4.6.2 NOE Restraints

For a structural model to be calculated from a protein assigned by NMR, distance restraints are required to guide the structure modelling algorithms. To generate distance restraints for His-RgpB-CTD, ^{15}N-HSQC-NOESY and ^{13}C-HSQC-NOESY experiments were recorded. The spectra derived from these experiments provide NOE through-space connections from the ^{15}N-^{1}H and ^{13}C-^{1}H spin systems, respectively. In these spectra, a proton resonance plus either a carbon or nitrogen resonance is present as the diagonal peak with all NOE connections of that atom appearing as cross-peaks. Cross-peaks in the F1 dimension are picked for each strip and assigned to the given spin system. However, the F3 dimension, that is the through-space coupling, is left ambiguously assigned. This allows the algorithmic program used in structural modelling, Aria 2.3 [152], more freedom to test different assignments to find the correct solution. Example strips for both ^{15}N and ^{13}C NOESY spectra, propagating from a single diagonal peak, are shown in Figure 4.6.1.

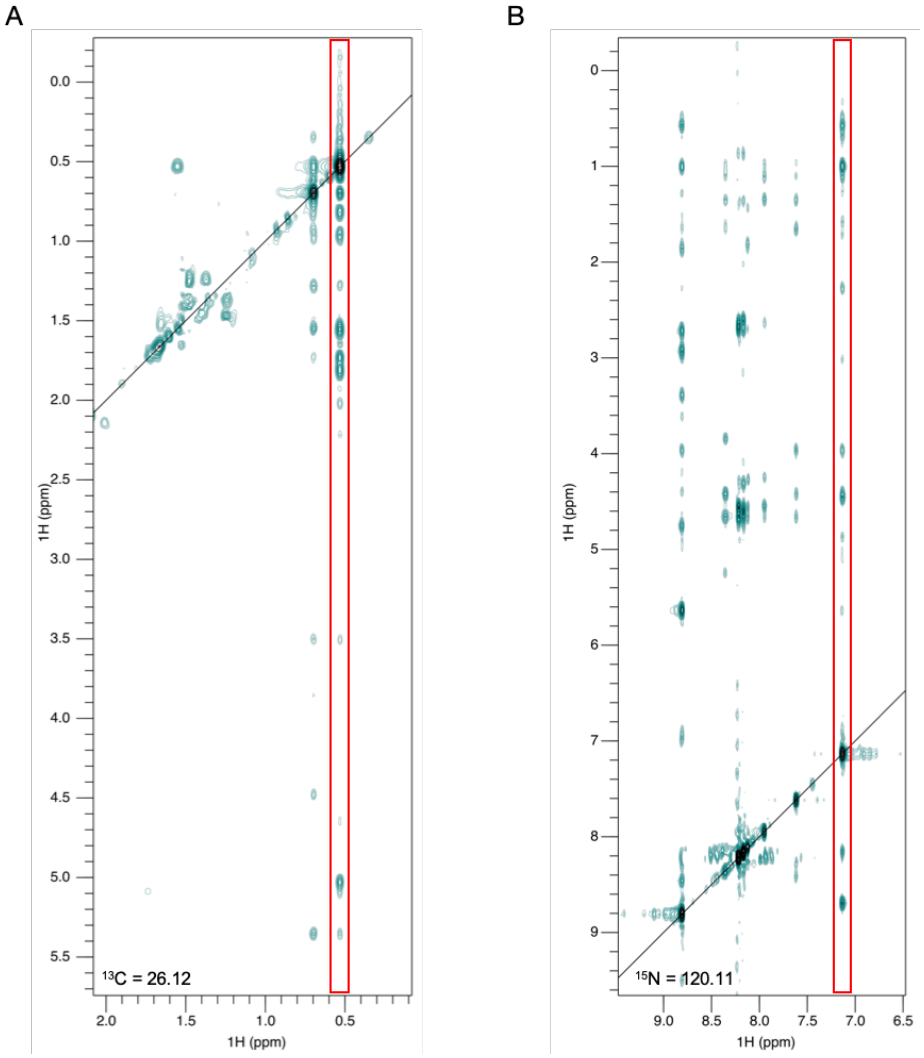


Figure 4.6.1: Example strips from ^{13} C-NOESY (A) and ^{15} N-NOESY (B) spectra for His-RpgB-CTD. Diagonal peaks are identified from previous assignments. Cross-peaks in the F1 dimension are picked for each strip, highlighted in red, and assigned to the given spin system. However, the F3 dimension, that is the through-space coupling, is left ambiguously assigned. The^{15}N-NOESY and ^{13}C-NOESY spectra were collected using a Bruker 800 MHz and 900 MHz NMR spectrometer, respectively, at 310K.

4.6.3 Hydrogen Bond Restraints

Hydrogen bond restraints were determined to further assist the structural calculation process. Because these bonds have a defined length they can be used to restrain two residues within 1.730 to 2.700 Å at the very start of the model refinement process. The 'long-range' HNCO is a three-dimensional experiment that shows the ^{15}N amide donor, $^{1}H_N$ and the ^{13}C' resonances [183]. That is, all partners in the hydrogen bond are visible and the restraint can be directly assigned using one peak. An example plane of the 'long-range' HNCO spectrum is shown in Figure 4.6.2. The through-bond HNCO spectrum is also displayed as it was used to assign the F3 (^{15}N) and F1 ($^{1}H^{N}$) resonances.

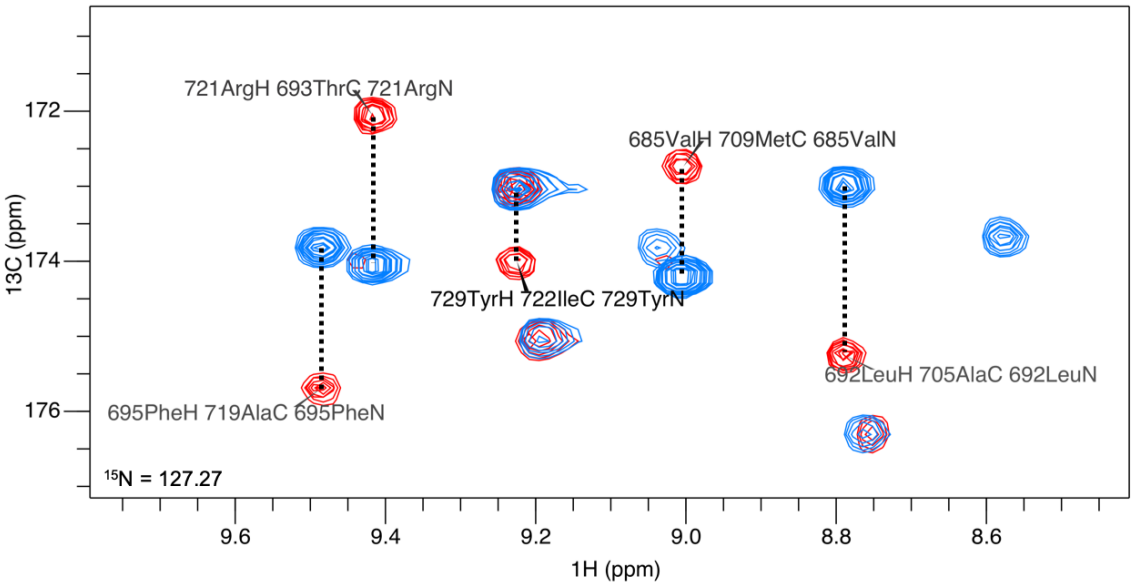


Figure 4.6.2: Long-range HNCO-COSY for identification of hydrogen bond restraints in His-RgpB-CTD. The HNCO spectrum, shown in blue, was used as a reference for F3 and F1 assignment. F2 assignment, for the C' hydrogen bond partner, was achieved using the red peaks in the long-range HNCO-COSY spectrum. The spectra were recorded using a Bruker 700 MHz NMR spectrometer at 310 K.

4.7 RgpB-CTD Structural Model

4.7.1 ARIA Structure Ensemble Modelling

Using the calculated restraint for His-RgpB-CTD it is possible to derive a structural model for the protein. Developing an accurate model of the protein tertiary structure depends on the accurate assignment of NOE cross-peaks as these confer information on the through-space distance between two proton nuclei in the system. Such an assignment can be performed manually but would be a very laborious task and could lead to the misassignment of connections. To perform this task in a faster and more accurate manner, structural calculations were performed using the ARIA 2.3 (Ambiguous Restraints for Iterative Assignment) software [152]. Providing ARIA with ambiguously assigned NOE peaks and a table chemical shifts for resonances in the protein, the program uses an iterative structural calculation scheme in which the full range of ambiguous distance restraints are tested. The calculation of the structures is driven using CNS (Crystallography & NMR System) [215] and ARIA analyses the conformers to filter out distance restraints that violate the structural arrangement. The structural calculation is also guided by the dihedral and the hydrogen bond restraints to further increase the accuracy of models generated. The total number of restraints that were committed to the final structural model ensemble are shown in Table 4.7.1.

4.7.2 The Solution Structure of RgpB-CTD

For the His-RgpB-CTD ensemble, 20 structures were calculated for each cycle with a total of eight iterations. The 10 lowest energy models from iteration eight were taken for further refinement by molecular dynamics in water. The His-RgpB-CTD ensemble, comprising the 10 lowest energy refined structures (Figure 4.7.1), shows a well-defined

Table 4.7.1: NMR assignment and structure statistics for RgpB-CTD.

NMR Distance Restraints	
Total Unambiguous Restraints	2207
Intra ($(\|i-j\|) = 0$ Å)	621
Sequential ($(\|i-j\|) = 1$ Å)	372
Short ($(\|i-j\|) < 3$ Å)	120
Medium ($(\|i-j\|) < 4$ Å)	60
Long ($(\|i-j\|) > 4$ Å)	510
Ambiguous Restraints	524
Total Dihedral Angle Constraints	92
Phi/Psi	46/46
Hydrogen Bond Restraints	22
Structural statistics	
Violations	
Distance restraints (Å)	0.00778 ± 0.00181
Dihedral angle restraints (°)	0.221 ± 0.0435
Deviation from idealised geometry	
Bond length (Å)	0.00316 ± 0.0000817
Bond angles (°)	0.412 ± 0.00928
Improper bond angles (°)	1.060 ± 0.0710
Average pairwise RMSD (Å)	
Backbone	0.311 ± 0.0239
Heavy atom	0.622 ± 0.0621
Ramachandran Statistics	
Favoured regions	88.9%
Allowed regions	10%
Disallowed regions*	1.1%

*All dihedrals in the disallowed region are in the disordered N-terminus and inter-strand loops.

backbone region for residues K^{672} - K^{736}, with the N-terminal residues G^{662} - D^{671} poorly defined. Per reside root mean square deviation (RMSD) values for the ensemble were calculated using CcpnNmr Analysis 2.4 [186]. Analysis calculates RMSD values by orientating models within the ensemble so that they are superposed. Structural variation and co-ordinate spread is measured for all residues in the protein sequence. Figure 4.7.2 shows the per residue RMSDs for all atoms and backbone atoms only. For the globular domain of the protein (K^{672} - K^{736}), the average backbone RMSD is 0.44 Å. A similar value was observed using average pairwise RMSD for heavy atoms and backbone atoms (Table 4.7.1). These scores demonstrate that there is a strong convergence of conformers in the final ensemble.

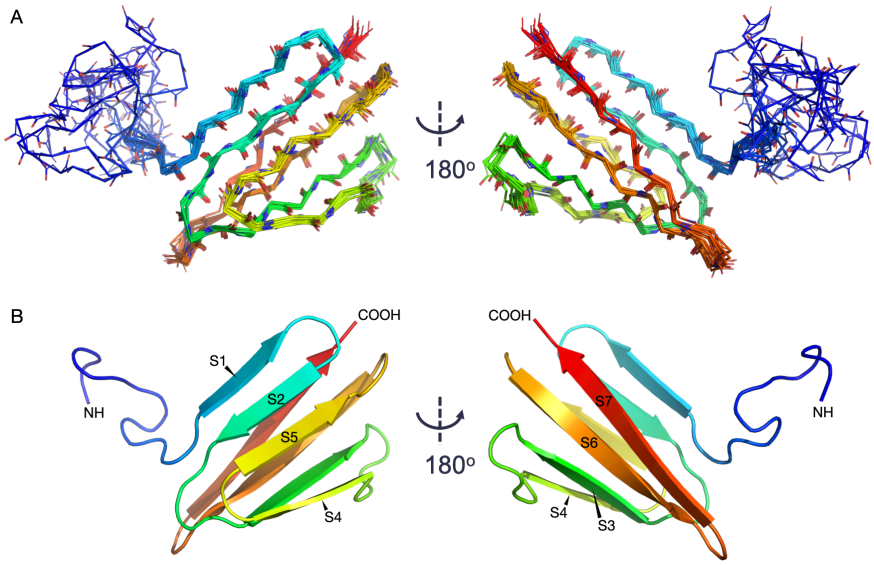


Figure 4.7.1: Structural ensemble of the 10 lowest energy water-refined models for His-RpgB-CTD. (A) The wireframe model shows the structural ensemble rotated by 180° in the y axis. (B) Cartoon model of the energy minimised average structure for RgpB-CTD rotated by 180° in the y axis, with β strands 1-7 highlighted. The N-terminus is shown in blue and the C-terminus is shown in red. His-tag residues not shown.

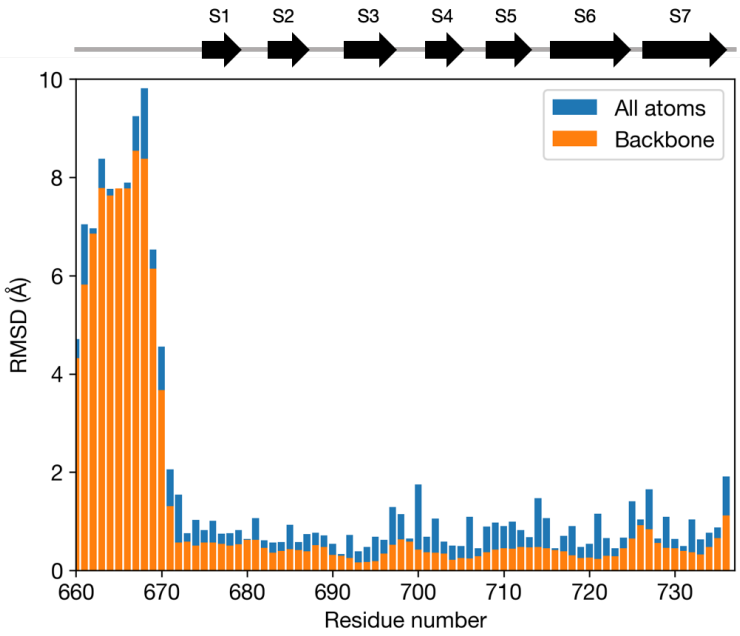


Figure 4.7.2: Per residue RMSD values for final 10 water refined His-RgpB-CTD structural models. Legend: Orange bars - backbone; Blue bars - all atom. All RMSD values were calculated using CcpnNmr Analysis 2.4.

4.7.3 Structure Model Validation

Table 4.7.1 lists the experimental restraint statistics used for structural model calculation. The listing of distance restraints represents all restraints that were used in the final ARIA run after optimisation and removal of violating restraints. The total number NOE restraints displayed represents merged NOEs merged from the ^{15}N and ^{13}C HSQC-NOESY spectra. For the unambiguous restraints, NOE distance is also listed.

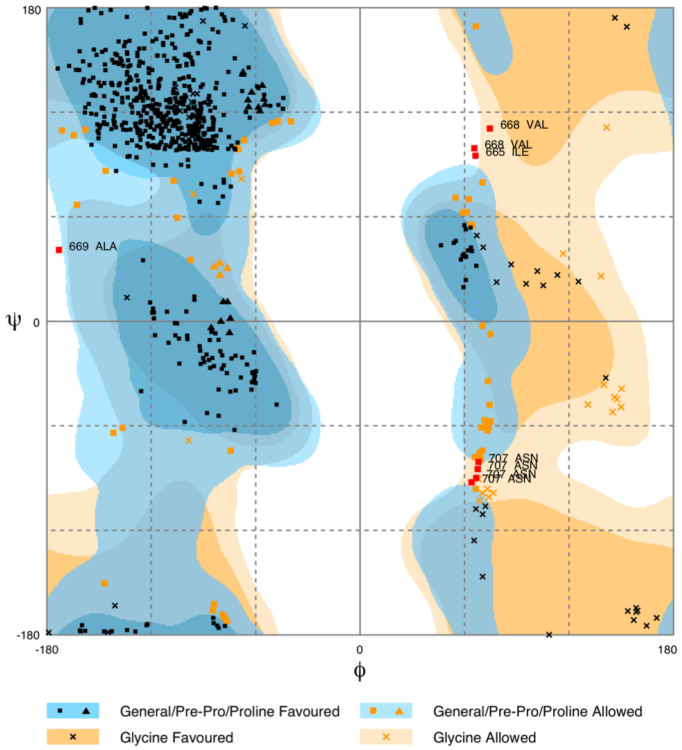


Figure 4.7.3: Ramachandran Plot for water refined structural ensemble of RgpB-CTD. Residues with dihedral angles in disallowed regions are highlighted in red. All disallowed dihedrals are for residues in the disordered N-terminal and loop regions.

The geometric quality of the His-RgpB-CTD structure was can be measured using the distribution of the backbone dihedral angles. Dihedral angle distribution was assessed using Procheck-NMR which analyses and compares the stereochemical geometry of ensemble structures [188]. The summary of Ramachandran statistics, assessed using the online RAMPAGE server [189], are shown in Table 4.7.1. 98.9% of residues have dihedrals within the favoured or allowed regions, indicative of a structure with good geometric quality (Figure 4.7.3). The remaining dihedrals in the disallowed region are from residues in the disordered N-terminus and loop regions. The 14 N-terminal residues, comprising the His-tag and enterokinase cleavage site, were not included in this analysis.

Restraint violations and deviations from idealised geometry refer to the 10 lowest

energy water-refined conformers. In addition to the average distance restraint violation score, no distance restraint violations above 0.5Å were observed. Average violations > 0.3Å and > 0.1Å were 0.2 ± 0.4 and 2.2 ± 1.6, respectively. All 92 DANGLE derived dihedral restraints and hydrogen bond restraints were committed to the final structure with no violations.

4.7.4 Structural Features of RgpB-CTD

As seen in Figure 4.7.1C the globular C-terminal domain (K^{672} - K^{736}) adopts an Ig-like fold comprised of seven antiparallel β-strands. The domain forms a compact β-sandwich fold with the S1, S2 and S5 strands tightly packed against the S3, S4, S6 and S7 β-strands. The S1 and S2 strands run along the N-terminal face of the sandwich, interspersed by a β-hairpin. A loop passes to the opposite, C-terminal face of the sandwich forming the S3 strand. S3 passes to S4 via a twisted beta-hairpin followed by a loop back to the N-terminal face. The S5 strand runs antiparallel to the S2 strand, succeeded by another loop to the C-terminal face where S6 and S7 complete the β-sandwich arrangement. The maximal dimensions of the β-sandwich are $23.8 \times 27.9 \times 37.5$Å, not including residues G^{662} - D^{671}.

The MUSCLE multiple sequence alignment for all W50 CTD proteins and the BlastP/Clustal Omega alignment for the W50 RgpB-CTD (Section 4.2.1) were uploaded to the online ConSurf server [166, 167] along with a representative RgpB-CTD structural model. The server was used to generate structural models coloured by sequence conservation (Figure 4.7.4B-C) and reveals that the most highly conserved residues are all located primarily at the C-terminal face of the domain. This region of conservation corresponds closely to the three conserved motifs previously identified by Veith *et al.*, [81] (Figure 4.7.4A). The S3-S4 loop in motif B is folded back with D^{696} forming hydrogen bonds to R^{700} and the highly conserved G^{699}. Motifs D and E com-

prise the S6 and S7 β-strands, respectively, which run anti-parallel to each other. This close spatial alignment of the three motifs adds further credibility to the hypothesis that these conserved regions play an important role in recognition by the T9SS.

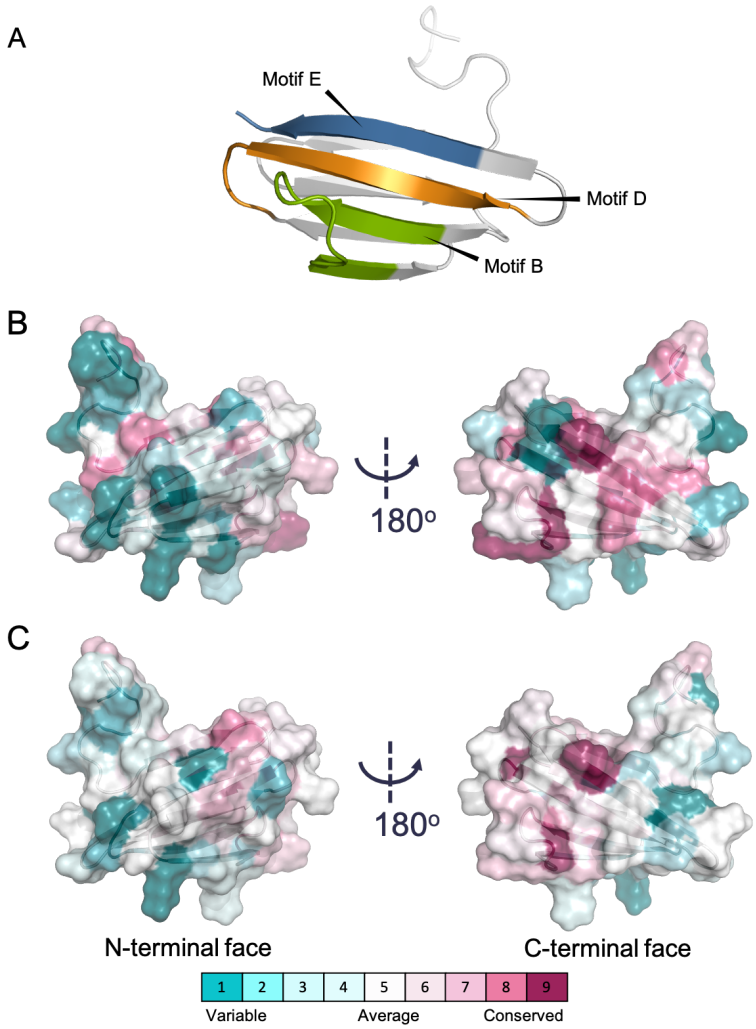


Figure 4.7.4: Spatial arrangement of conserved residues in RgpB-CTD. (A) Cartoon of the RgpB-CTD with the three conserved motifs identified by Veith *et al.*, [81] highlighted in blue, orange and green. (B) ConSurf colouring for RgpB-CTD using the BlastP/Clustal Omega alignment for RgpB-CTD. (C) ConSurf colouring for RgpB-CTD using the MUSCLE alignment of all W50 CTDs. A water refined ensemble average energy-minimised model of RgpB-CTD was used for all of the images.

4.8 RgpB-CTD Dynamics

4.8.1 Enterokinase Cleavage of His-tag

During the process of assigning the NMR spectra for RgpB-CTD, an X-Ray crystallographic structure was published for the same domain with the PDB code, 5HFS [170]. In this structure, the CTD is shown to form a dimer via domain swapping of the S7 strands. In addition, the authors demonstrated dimerisation by cross-linking with 0.1% glutaraldehyde. The construct used for these experiments contained residues K^{672} - K^{736} expressed with an N-terminal GST tag. This tag was removed prior to crystallisation and cross-linking studies.

The formation of a dimer is not something that was observed during sample preparation and NMR data collection as part of this study. This conclusion is based on the elution volume during size exclusion chromatography and peak line-widths during NMR data acquisition. To test the formation of an RgpB-CTD dimer the cross-linking experiments performed by de Diego *et al.*, [170] were repeated. To ensure there was no interference from the N-terminal His-tag it was removed by enterokinase cleavage.

Pooled sample from Ni-NTA purification was used directly for enterokinase cleavage. The sample was transferred to enterokinase cleavage buffer, via PD-10 buffer exchange, followed by cleavage overnight at 25 °C. The processed sample was then purified by reverse Ni-NTA chromatography, using a gravity flow column, to remove enterokinase and uncleaved sample. Flow-through and wash from reverse Ni-NTA chromatography was pooled and concentrated. This was loaded onto the HiLoad 16/600 Superdex 75 pg column to exchange buffer and remove any remaining contaminants. Figure 4.8.1A shows that RgpB-CTD elutes as a single major peak. Fractions of this peak show it is a single clean band of the correct size with only minimal degradation (Figure 4.8.1B).

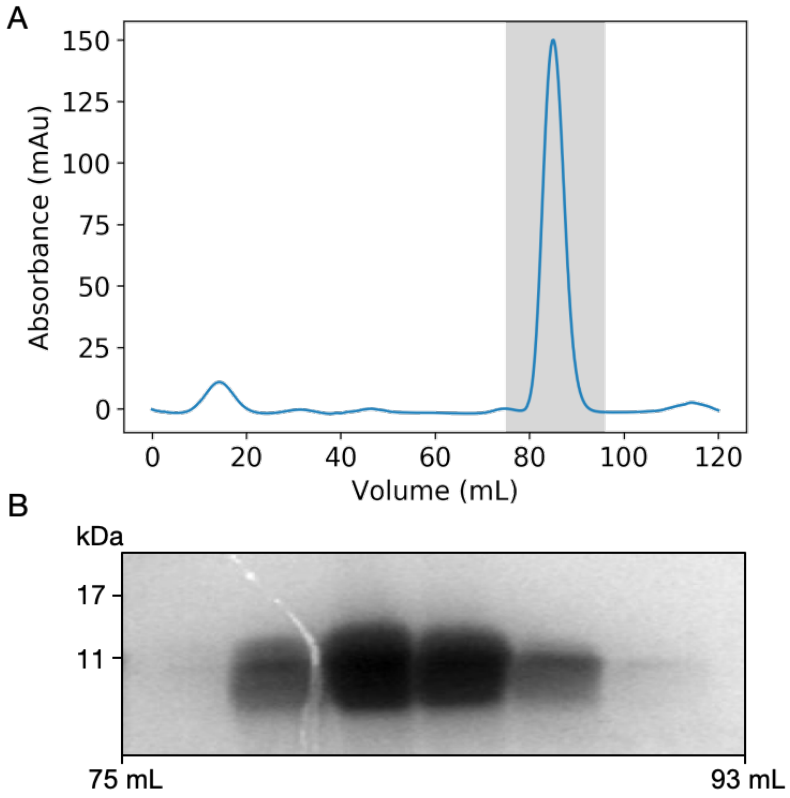


Figure 4.8.1: Result of size exclusion chromatography for RgpB-CTD following enterokinase cleavage. (A) UV A280 absorbance chromatogram with the RgpB-CTD peak highlighted in grey. (B) SDS-PAGE of the RpgB-CTD 3 mL fractions from 72 mL to 90 mL highlighted in grey on the above chromatogram.

4.8.2 RgpB-CTD Glutaraldehyde Cross-linking

To test the formation of an RgpB-CTD dimer the glutaraldehyde cross-linking experiment was repeated using 0.1% glutaraldehyde and increasing concentration of RgpB-CTD (0.01, 0.5 and 1 mg mL^{-1}). The same method as described by de Diego *et al.*, [170] was used for the reaction. However, instead of observing two bands, representing a monomer and dimer, multiple oligomerisation states could be observed (Figure 4.8.2).

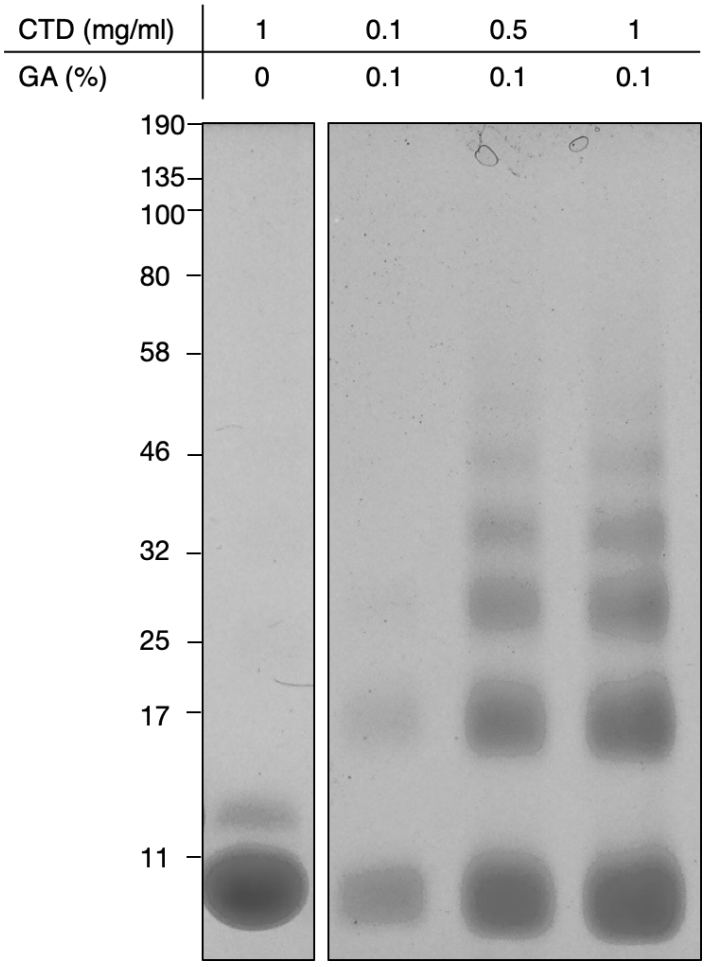


Figure 4.8.2: RgpB-CTD oligomerisation analysis by Glutaraldehyde cross-linking. Multiple oligomerisation states can be observed for RgpB-CTD following the addition of glutaraldehyde.

4.8.3 T1 and T2 Relaxation

Due to the inconclusive nature of the cross-linking experiment, NMR relaxation experiments were collected for RgpB-CTD to determine the molecular dynamics of the protein domain. The aim of these experiments was to ascertain if there is increased flexibility in the S7 strand, indicative of a potential domain swapping event. NMR relaxation times can also be used to calculate the correlation time of a protein which can be used to estimate protein molecular weight.

T1 and T2 relaxation data were analysed using the ^{15}N-HSQC spectrum as a

reference for peak assignments. In order to reduce the overlapping of peaks in the middle of the spectrum, relaxation data was collected for RgpB-CTD with the His-tag removed. Comparison of peak locations to the ^{15}N-HSQC spectrum for His-RgpB-RCTD shows that cleavage of the His-tag does not significantly alter the peak locations 4.8.3.

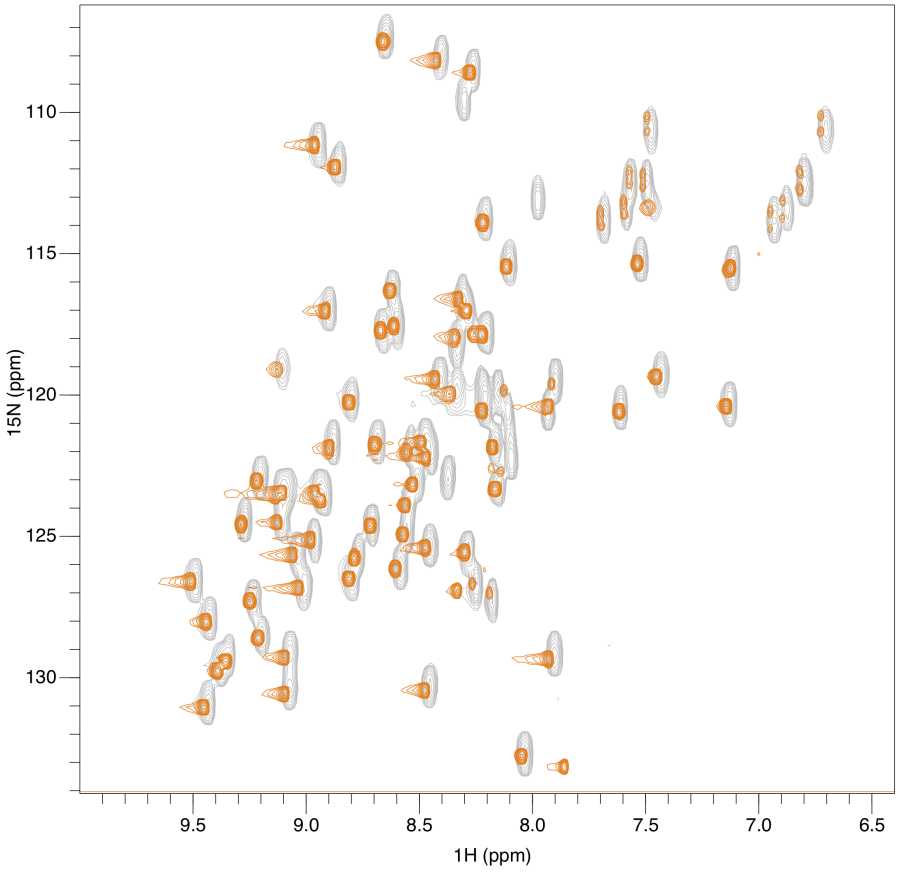


Figure 4.8.3: Comparison of ^{15}N HSQC spectra for RgpB-CTD with and without the N-terminal His-tag. The spectrum for His-RgpB-CTD is shown in grey and the spectrum for RgpB-CTD is shown in orange. Other than the loss of peaks, due to removal of the 13 N-terminal protein residues, only minor changes in peak location can be observed and these are limited to the C-terminus.

Relaxation rates were calculated for T1 and T2 spectra using the exponential function $A\,exp(-Bx)$. Peaks with poor curve fitting, due to peak overlap, were removed and not analysed. The average T1 and T2 relaxation times for the structured region

of the protein (K^{672} - K^{736}) were calculated to be 545.38 ± 80.46 ms and 218.11 ± 63.58 ms, respectively. The T1 and T2 times are plotted against residue number in Figure 4.8.4. Analysis shows a uniform pattern for most residues in the β-sandwich domain. An increased relaxation time was observed for the flexible N-terminus as expected. The T1 T2 ratio in the s7 β-strand, which is meant to be domain swapping to form dimers, shows the same level of structural dynamics as the other β-strands in the CTD.

The average correlation time for the domain was calculated, using Equation 4.8.1 derived from eq 8. in Kay *et al.* [216], to be $\tau_c \sim 3.3$ ns [216]. The online Protein Correlation Time Calculator tool [217] was used to predict the expected correlation time for the RgpB-CTD monomer at 310K. The predicted value was found to be between 3.1-3.8 ns suggesting that the RgpB-CTD observed here is in a monomeric state.

$$\tau_c \sim \frac{1}{4\pi\upsilon_N}\sqrt{6\frac{T_1}{T_2} - 7} \tag{4.8.1}$$

where υ_N is the resonance frequency of ^{15}N nuclei in Hz in 700mHz *Bo* field.

4.8.4 Heteronuclear NOE

^{1}H-^{15}N heteronuclear NOE (hetNOE) experiments were recorded alongside relaxation experiments to provide supportive information on the N-H bond motions of residues in the RgpB-CTD protein. Two hetNOE spectra were recorded, the first with proton saturation and the second without. Residues with faster motion, indicative of increased flexibility in the protein backbone, show a decreased NOE intensity relative to other residues in the protein [218]. Values of steady-state NOEs were calculated from the ratio of peak intensities (saturated over unsaturated).

The spectrum derived from these experiments is comparable to an ^{15}N-HSQC spec-

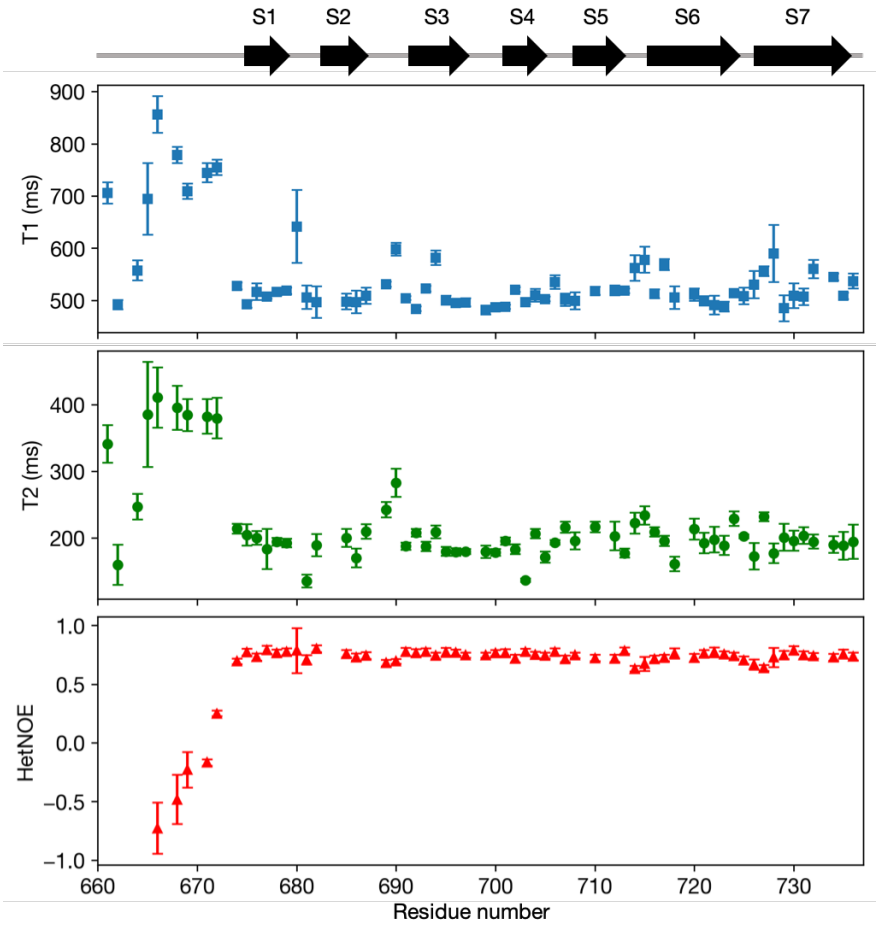


Figure 4.8.4: HetNOE, T1 and T2 relaxation times for each residue in RgpB-CTD. Increased flexibility can be seen for residue in the disordered N-terminus do not within the β-sandwich fold of the CTD.

trum, allowing easy propagation of peak assignments from the HSQC to HetNOE spectra. The HetNOE experiments were recorded for RgpB-CTD with the His-tag removed. To ensure this had no impact on the location of other peaks in the spectra, the Heteronuclear NOE spectra were compared to the ^{15}N-HSQC containing the His-tag. As shown in Figure 4.8.5 no major changes to peak location could be observed beyond the loss of peaks due to the removal of His-tag residues. Peaks were selected using the ^{15}N-HSQC peaks as a reference with assignment propagation being possible for all residues. Residues with overlapping peaks were discarded from analysis due to inaccuracies in measuring intensity difference.

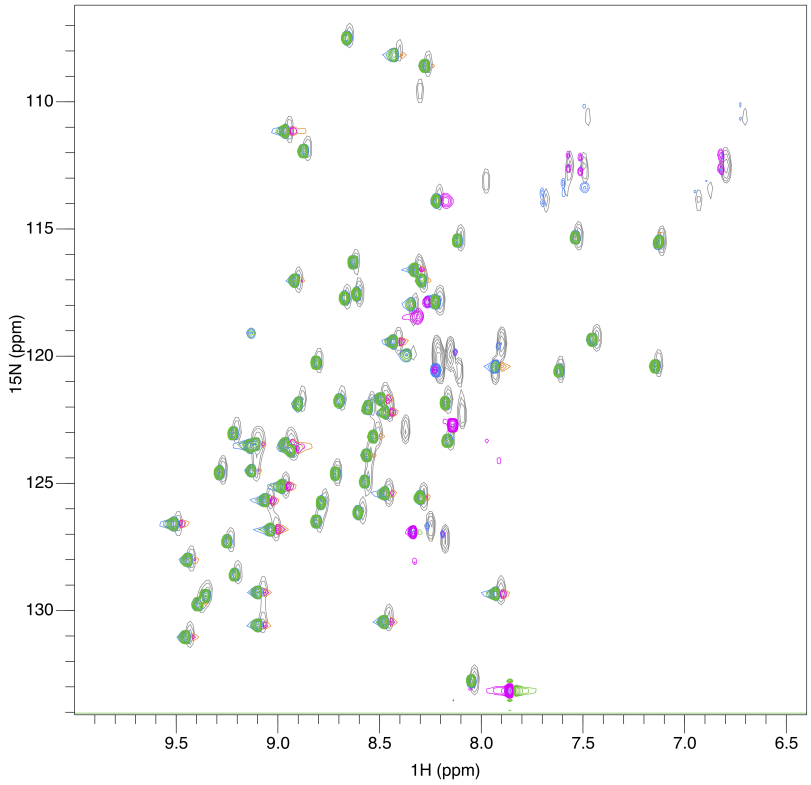


Figure 4.8.5: Heteronuclear NOE spectra for RpgB-CTD compared to the ^{15}N-HSQC for His-RgpB-CTD. The saturated HetNOE spectrum is shown in green, positive peaks, and pink, negative peaks. The unsaturated HetNOE spectrum is shown in blue, positive peaks, and orange, negative peaks. The ^{15}N-HSQC for His-RgpB-CTD, used as a reference for peak assignment, is shown in grey.

The HetNOE ratios are shown in Figure 4.8.4 and demonstrate that increased motion, relative to the folded protein domain, is seen for the flexible N-terminus (G^{662} - K^{672}). The average Het-NOE score for the full protein was calculated to be 0.656 ± 0.300. The average score for the folded β-sandwich domain only has an average score of 0.740 ± 0.0369. No deviation from this average was observed in the S7 strand involved in dimer formation observed in the crystallographic model.

4.9 Discussion

When work for this chapter was initiated no structural models were available for the CTD of any protein secreted by the T9SS. In this chapter, the CTD from RgpB was successfully cloned and expressed recombinantly in *E. coli*. The protein was purified and NMR experiments used to assign chemical shifts values for nuclei in the protein. Structural modelling using distance restraints and chemical shift values derived from the NMR experiments revealed that it possesses and an Ig-like fold comprised of seven anti-parallel β-strands. This structural motif is widely observed in a diverse set of protein families isolated from eukaryotes, prokaryotes and archaea [219]. The domain is involved in a variety of functions often mediated by protein-protein interactions - consistent with the proposed function of the T9SS cargo CTD [220].

The same fold was observed for the crystallographic model of RgpB-CTD, which also contained the preceding IgSF domain of RgpB [170]. Alignment of the average energy minimised NMR model to the IgSF-CTD crystal model using the online server TM-Align [221] reveals that the models have an RMSD of 0.74 Å and a TM score of 0.82 (Figure 4.9.2). Although the models are well aligned the packing of the IgSF domain against the CTD does not allow for the linker region to connect the two domain. This suggests that the packing of the two domains may be due to crystallisation conditions and the CTD may orient differently relative to the IG-SF domain in solution state.

When RgpB-CTD was crystallised without the IG-SF domain the resulting structural model was dimeric via domain swapping of the C-terminal S7 strand (Figure 4.9.1)[170]. Given the essential function of the S7 strand in recognition by the T9SS [97], the formation of CTD dimers via domain swapping of this strand would have significant implications for cargo protein recognition. However, oligomerisation of the RgpB-CTD construct used for the NMR experiments described above was not observed, nor was there any indication of chemical exchange between a dimeric and

monomeric state. Furthermore, correlation time calculated from T1 and T2 relaxation data for the Rgp-CTD are consistent with a monomeric state of the domain. It has also been demonstrated for a range of proteins that individual domains may form non-functional domain-swapped dimers when crystallised, whereas the full-length protein does not [222]. It is therefore possible, that the domain swap observed for the crystal model of RgpB-CTD is an artificial state induced by non-physiological crystal packing conditions.

Another possible explanation of the dimerisation is that the RgpB-CTD construct expressed for crystallisation possessed a GST purification tag. The GST tag is itself a dimeric protein and has previously been shown to induce the formation of dimers in fusion proteins [223, 224]. Bringing two CTD molecules together via the addition of the GST tag may lead to an increased propensity for domain swapping and the formation of dimers.

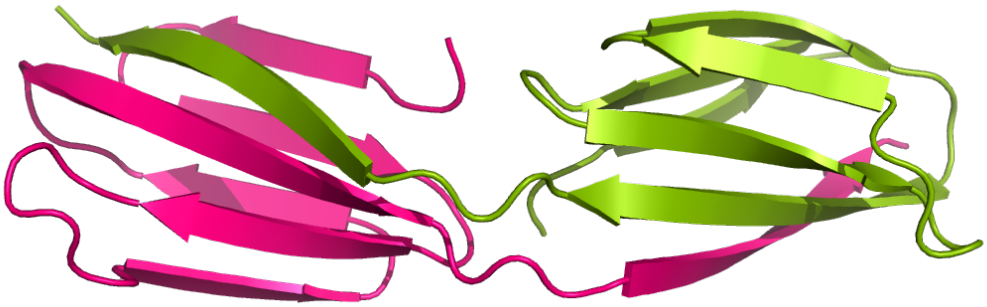


Figure 4.9.1: Dimerisation of RgpB-CTD via domain swapping of S7 strand. The crystal structure of RgpB-CTD (PDB: 5HFS) shows the protein in a dimeric state formed by domain swapping of the S7-strand. However, no dimerisation is observed in the IgSF-CTD crystal model (PDB: 5AG8) or the NMR structural models presented here.

Glutaraldehyde cross-linking was used by de Diego *et al.*, [170] to show the formation of CTD dimers. The CTD construct used here, which contains addition N-terminal residues G^{662} - D^{671}, revealed the formation of multiple oligomerisation states and visible laddering of the protein in an SDS-PAGE gel (Figure 4.8.2). Glutaralde-

hyde forms cross-links predominantly between the ϵ-amine group of lysine side-chains which can trap proteins in complexes or oligomeric states. Due to the high number of surface-exposed lysines in RgpB-CTD, the multiple oligomeric states observed when using glutaraldehyde may be due to non-specific cross-linking of lysines rather than from multiple domain swap events.

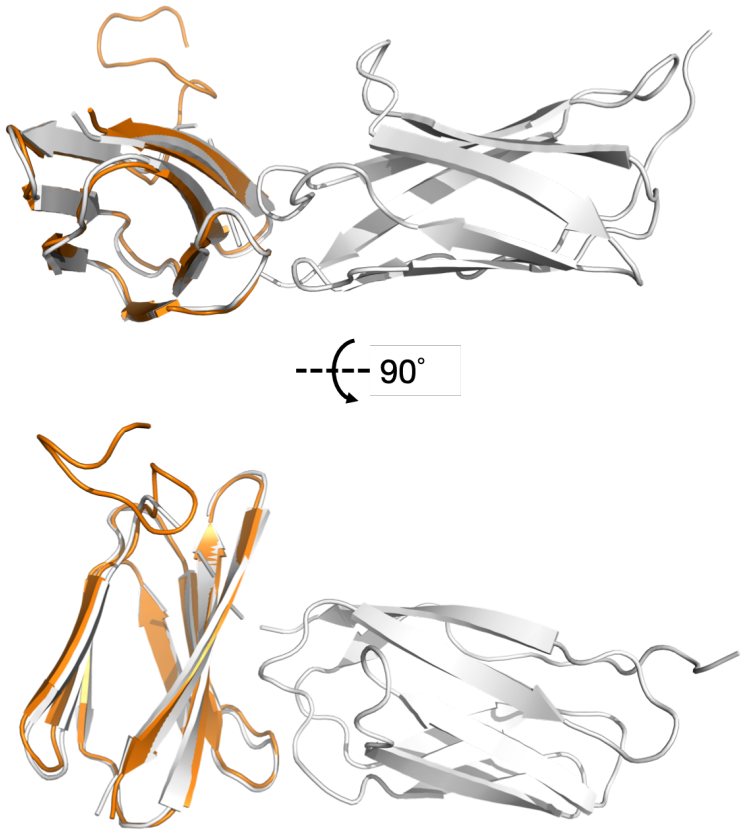


Figure 4.9.2: Comparison of RgpB-CTD NMR and crystal structures. Alignment of the NMR model (orange) to the crystal model (grey) for RgpB.

Structure based alignment of the NMR structural model to the recently resolved crystal models for two other T9SS cargo protein CTDs (PorZ [92] and Hbp35 [225]) reveals that the Ig-like fold is well conserved (Figure 4.9.3). This is despite having a relatively low sequence identity: 24.3% between RgpB and PorZ and 27.5% between RgpB and Hbp35. This adds support to the hypothesis that structural conservation within the CTD, and not only the conservation of sequence motifs, is important for

the function of the domain. Comparison of the three available structural models was performed using the TMAlign web server [221]. This server calculates a TM-score for the alignment which is more sensitive to global fold similarity than RMSD scoring, with an expected score of < 0.17 for two randomly aligned proteins with a score of 0.5-1.0 for proteins that possess the same fold according to SCOP and CATH classifications [221]. This analysis revealed a TM-score of 0.76 for RgpB-CTD compared to PorZ-CTD and and a TM-score of 0.69 for RgpB-CTD aligned with Hbp35-CTD. This shows that the Ig-like fold is conserved between the known CTD structures of T9SS cargo proteins, despite the low level of sequence identity.

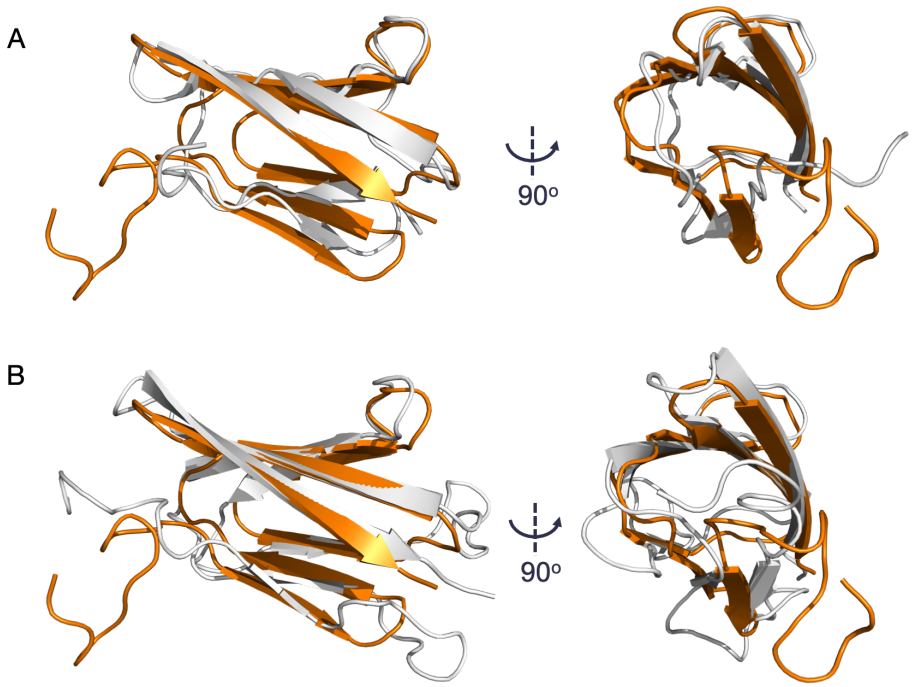


Figure 4.9.3: Comparison to RgpB, PorZ and Hbp35 CTDs. (A) Comparison of averaged NMR model of RgpB-CTD (orange) against the crystal model of Hbp35-CTD (grey). (B) Comparison of averaged NMR model of RgpB-CTD (orange) against the crystal model of PorZ-CTD. Alignments were performed using the TMAlign web server [221].

Superposition of the RgpB, Hbp35 and PorZ models, along with ConSurf analysis for RgpB, reveals that the C-terminal face of the domain is the most highly conserved

region. This corresponds closely to the three conserved motifs B, D and E previously identified by Seers *et al.*, [80]. Several of the most highly conserved residues, located in motifs D and E, correspond perfectly to strands S6 and S7, previously demonstrated to be sufficient for export of GFP [97]. The high degree of structural similarity within the S6 and S7 strands, and the presence of the highly conserved residues at the C-terminal face, gives support to the idea that the C-terminal face of the domain is important for the recognition of cargo proteins by the T9SS.

Truncation of the two, four and ten C-terminal residues of RgpB cause the protein to accumulate in the periplasm and prevent protein secretion [226]. It has therefore been suggested that these residues are essential for cargo protein recognition. However, given the β-sandwich fold observed here it is possible that the truncation of these residues prevents the correct folding of this domain.

There is one example in the literature of a point mutation in the RgpB-CTD the fully prevents the surface attachment RgpB [207]. The $D^{696}A$ mutant showed a strong band corresponding to unprocessed RgpB in the clarified culture fluid and no periplasmic accumulation. It was therefore suggested, that this residue, located in conserved motif B, is important for processing and surface attachment but not for secretion. Analysis of D^{696}s in the NMR model of RgpB-CTD shows it forms a hydrogen bond to the highly conserved G^{699} residue and to R^{700}. The D^{696} side-chain may help to stabilise the structural motif in this loop facilitating cargo protein processing. However, the same motif can be observed in the structural model of PorZ which does not undergo surface attachment. It is interesting to note that the corresponding residue in Hbp35-CTD is asparagine but the structural motif is still maintained. It remains unclear how this motif may direct cargo protein attachment. Given that the $D^{696}A$ mutant localises to the clarified culture fluid, this motif may be important in facilitating cell surface attachment of cargo proteins prior to LPS

modification.

4.10 Summary

This chapter describes how a representative CTD was selected for *P. gingivalis* T9SS cargo proteins and the process by which it was cloned, expressed and purified. NMR spectroscopy was used to generate chemical shift values and distance restraints for the majority of atoms in this domain. These were used to derive a structural model of the domain revealing that it has an Ig-like fold comprised of seven β-strands. An X-ray crystallographic model of this domain has since been published for this domain suggesting it is dimeric. To test this NMR relaxation experiments were performed to determine flexibility in the C-terminal β-strand and to estimate the molecular mass of the protein construct. The data presented here suggested that the CTD is in a monomeric state in solution and dimerisation may be a result of the crystallisation process. Comparison of the RgpB-CTD to other more recently published structures of the PorZ and Hbp35 CTDs reveal that this fold is well conserved between cargo proteins despite a low level of sequence identity. This further supports the hypothesis that structural motifs are important for cargo protein recognition by the T9SS.

Chapter 5

Screening for Interactions Between RgpB-CTD and the T9SS

5.1 Chapter Aims

The aim of this chapter was to identify components of the T9SS from P. gingivalis capable of interacting with RgpB-CTD. Due to the relative novelty and lack of detailed characterisation for the T9SS, a non-specific approach for interaction screening was employed. This was to ensure known system components with unknown function or novel system components could be identified. The primary purpose of doing this was to develop a pool of potential interaction partners that could be tested in more detail, potentially including structural studies of an interaction complex. However, the aim was not to perform an exhaustive screen for the identification of all potential interactions. Instead, the aim was to immediately take forward the most promising candidate protein(s) for interaction verification and further characterisation.

5.2 His-RgpB-CTD:*P. gingivalis* Lysate Pull-Down

Pull-down assays are commonly used *in vitro* techniques for the identification of unknown protein-protein interactions. The technique is based on affinity chromatography, in which the bait protein is captured by an immobilised affinity ligand. The immobilised bait protein is then incubated with cell lysate containing the potential interaction partners. Following a series of washing steps, to remove non-interacting proteins, the sample is eluted to release the bait protein and any captured prey proteins.

In this study, the bait protein used was the His-tagged RgpB-CTD construct described in Chapter 4. The *P. gingivalis* strain W50 (ATCC 53978) was selected as a source of prey proteins for the same reasons as described in Section 4.2.2. Again, in brief, this strain is one of clinical significance and has the widest available array of deletion mutant constructs for additional analysis. Ni-NTA agarose was used for immobilisation of His-RgpB-CTD via His-tag affinity binding. 10 mL cultures of *P. gingivalis*, grown for 24-hours in BHI-H broth, was collected and lysed by sonication. This lysate was added to the resin bed and incubated for 1 hour at 4 °C. The sample was then washed with Ni-NTA wash buffer containing detergent to maintain solubility of membrane proteins (20 mM Tris, pH 8; 200 mM NaCl; 10 mM imidazole; 0.1% Triton-X100). The sample was washed for 3×15 minutes at 4 °C, using 5 column volumes of wash buffer. Samples were eluted directly into SDS-PAGE loading buffer and heated to 70 °C for 20 minutes. This method of elution was used to ensure that all protein sample captured by the resin, or the CTD bait protein, were released to allow for identification in subsequent SDS-PAGE analysis.

Figure 5.2.1 shows the gel image following SDS-PAGE separation of eluted pull-down samples. His-RgpB-CTD, with no W50 lysate washed over, and W50 lysate washed over resin, with no captured bait protein, were used as negative controls. Many

non-specific bands could be identified in the W50 only sample, suggesting that proteins from the cell lysate are able to interact with the Ni-NTA resin. Some non-specific bands could also be seen in the His-RgpB-CTD only lane as well. The test sample, containing His-RgpB-CTD and W50 lysate showed no additional bands relative to the His-RgpB-CTD only negative control.

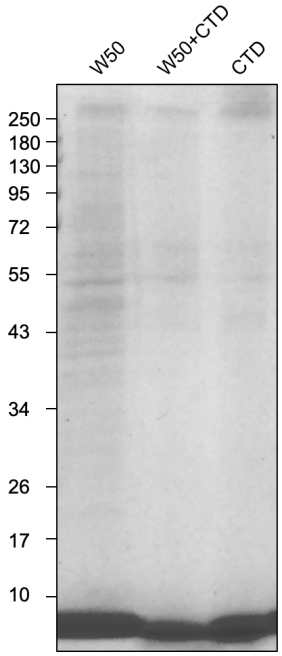


Figure 5.2.1: SDS-PAGE analysis of His-RgpB-CTD pull-down against *P. gingivalis* lysate. No additional bands can be identified in the test sample (W50 + His-RgpB-CTD) compared to the two control samples. Number markers on the left-hand side represent the size of marker proteins in kDa. Legend: W50 - *P. gingivalis* W50 lysate, CTD - His-RgpB-CTD.

5.3 Formaldehyde Cross-linking of His-RgpB-CTD: *P. gingivalis* Interactions

A caveat with the technique described above is that during sample washing transiently interacting proteins will be washed off. Therefore, the method was altered to

incorporate chemical cross-linking. By introducing chemical cross-links between the two binding partners it is possible to capture more transient protein interactions [**?**]. In addition, more stringent washing conditions can be used without losing captured protein complexes. The washing step was increased from three 15 minutes washes to three 30 minutes washes to remove non-specific bands in the W50 lysate only and His-RgpB-CTD only control samples.

Formaldehyde was selected as a chemical cross-linker for this experiment as it is a non-specific cross-linker. Therefore, any protein complex can theoretically be captured using formaldehyde cross-linking. Formaldehyde is a small electrophilic molecule susceptible to nucleophilic attack by amino acids [227]. Formaldehyde is initially attacked by a primary amine in the protein and leads to the formation of a Schiff base. The Schiff base is then susceptible to nucleophilic attack by another nucleophile. This could be another protein brought into proximity of the Schiff base by interaction with the initial protein. The second nucleophilic attack causes the formation of a methylene bridge that covalently links the two proteins [228]. This bridge links nucleophilic groups of proteins that are 2 Å apart, making it well suited to capturing interacting proteins and reduces the number of non-specific interactions that would be captured by use of chemical cross-linkers with longer linkage distances.

The same pull-down method was used as before, however, His-RgpB-CTD was added directly to W50 lysate with either 0.1 or 1% formaldehyde. PBS was used for cross-linking reactions as it is an amine-free buffering agent, unlike Tris, which does not react with formaldehyde. The cross-linking reaction was performed at 37 °C for 30 minutes followed by the addition of glycine to a final concentration of 125 mM to quench the reaction. Following 15 minutes of quenching at room temperature 8M urea and 1% SDS was added to the cross-linked samples to denature proteins in the sample. The suspension was loaded onto Ni-NTA resin and washed with Ni-NTA

wash buffer supplemented with 8M urea and 1% SDS. Samples were eluted directly into SDS-PAGE loading buffer and heated to 70 °C for 20 minutes.

Recombinantly expressed His-tagged LspD-N0 protein, from *Legionella pneumophila*, was used as an additional control to test for non-specific binding. LspD-N0 is the N0 domain taken from the Type 2 Secretion System secretin LspD [229]. This protein was selected because it is of a similar size to His-RgpB-CTD and because it also possesses an N-terminal His-tag. This construct was readily available in our lab and has no known or expected interactions with the T9SS. The W50 lysate with His-RgpB-CTD and no formaldehyde cross-linking sample was repeated to use as a negative control for the pull-down. W50 lysate and His-RgpB-CTD were each incubated with formaldehyde alone to provide further negative controls for the experiment.

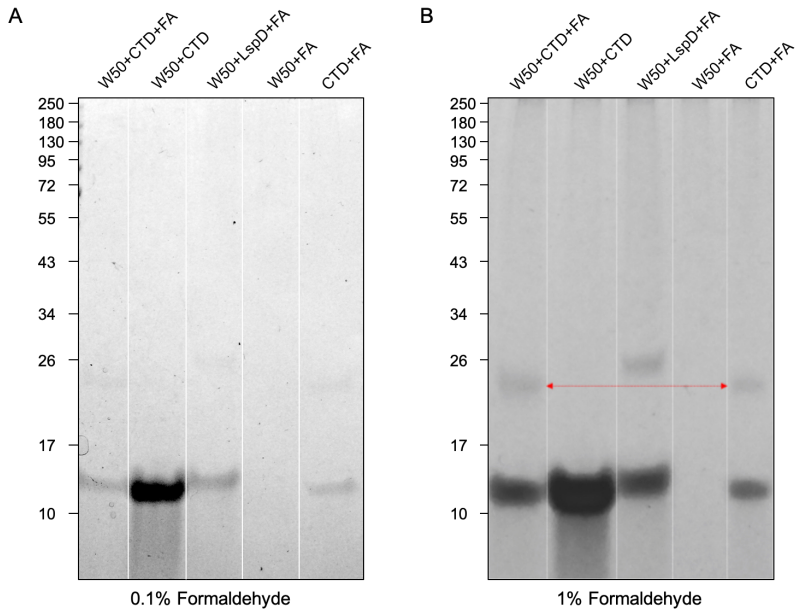


Figure 5.3.1: SDS-PAGE analysis of His-RgpB-CTD pull-down against *P. gingivalis* lysate with formaldehyde treatment. Cross-linking of pull-down samples was performed using 0.1% (A) and 1% (B) formaldehyde. No additional bands can be observed in the test samples relative to the control. The higher band, indicated by the red line, represents the cross-linking induced oligomerisation of His-RgpB-CTD. Legend: W50 - *P. gingivalis* W50 lysate, CTD - His-RgpB-CTD, FA - formaldehyde.

Figure 5.3.1 shows the result of SDS-PAGE separation of formaldehyde cross-linked pull-down samples. Only one higher band can be observed in the test sample (W50+CTD+FA). However, this band is also present for His-RgpB-CTD incubated alone with formaldehyde. The size of this band suggests it corresponds to the CTD dimer that is formed by the addition of a chemical cross-linker, as described in chapter 4. An additional band can also be seen for LspD-N0. This band corresponds to the dimeric form of LspD, consistent with previous observations from our lab.

5.4 Recombinant Expression of His-Tagged CTD in *P. gingivalis*

5.4.1 Construct Design and Assembly

Because of the difficultly in identifying protein complexes via the addition of His-RgpB-CTD to *P. gingivalis* lysate it was reasoned that the CTD may only interact with an intact secretion system and that capturing complexes *in vivo* would prove more effective in identifying interaction partners. To pursue this hypothesis, constructs were designed for His-tagged CTD from RgpB using pTIO-AI as a vector. The pTIO-AI vector was gifted by Dr. Mark Roberts and is derived from the vector pTIO-1 by the addition of the RgpA promoter to allow recombinant expression of protein constructs within *P. gingivalis* [230, 231]. This vector possesses an Amp^R cassette to allow positive selection in *E. coli* and possesses and ErmF cassette to allow positive selection in *P. gingivalis*.

Constructs were designed using Gibson Assembly to allow the single-step assembly of multiple fragments by incorporating the necessary overhangs into primer sequences. Constructs of RgpB-CTD were designed to carry a 6×His-tag on either the N- or C-terminus. Both N- and C-terminally tagged constructs were created to mitigate

against any potential interference in protein interactions arising from the addition of artificial protein sequences. However, it has previously been demonstrated that the addition of a 6×His-tag to the C-terminus of PorZ does not disrupt export of this protein to the outer membrane [92].

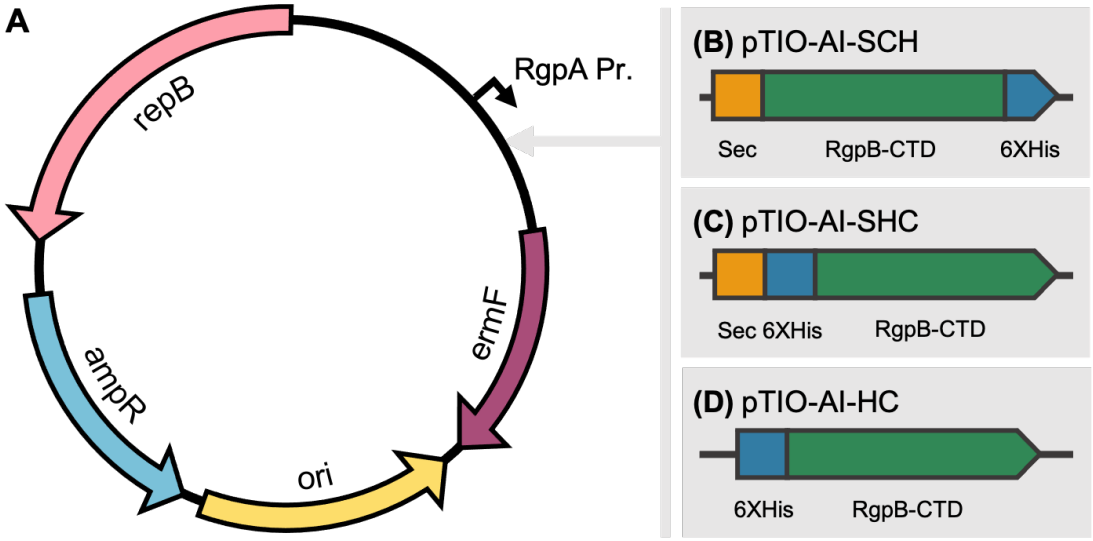


Figure 5.4.1: Plasmid maps for pTIO-AI constructs containing RgpB-CTD variants. (A) circular plasmid map of pTIO-AI prior to assembly. (B) Linear map of pTIO-AI-SCH containing the Sec signal peptide, RgpB-CTD with a C-terminal 6X His-tag. (C) pTIO-AI-SCH containing the Sec signal peptide, RgpB-CTD with an N-terminal 6X His-tag. (D) Linear maps of pTIO-AI-SHC containing RgpB-CTD with an N-terminal 6X His-tag.

The N-terminal signalling peptide of RgpB, for Sec mediated trafficking to the periplasm, was added to the N-terminus of these constructs. An additional construct, containing a C-terminal His-tag but no Sec signalling peptide was created to use as a negative control. A plasmid map for the three constructs can be seen in Figure 5.4.1. The constructs were named pTIO-AI-CH, pTIO-AI-SCH and pTIO-AI-SHC according to the arrangement of Sec signalling sequence (S), His-tag (H) and CTD (C) fragments.

The NcoI and BamHI cut-sites in pTIO-AI, downstream of the RgpA promoter, were used to linearise the vector for fragment insertions. Following assembly of the

pTIO-AI constructs using the NEBuilder® HiFi DNA Assembly Cloning Kit, the sample was transferred to *E. coli* NEB5α cells by heat-shock transformation. Positive colonies identified on plates containing ampicillin were used for plasmid mini-prep and insert sequences were confirmed by Sanger sequencing (SourceBioscience).

Following sequencing confirmation of plasmid assembly, the pTIO-AI constructs were electroporated into *P. gingivalis* W50 with positive selection using erythromycin. Using this method erythromycin-resistant strains were successfully established for each of the CTD plasmid constructs.

5.4.2 Periplasmic Extraction of His-tagged CTD

To check that the pTIO-AI constructs were successfully expressing folded protein and that trafficked to the periplasm, a periplasmic extraction followed by nickel affinity chromatography was performed for each construct. 10 mL cultures of *P. gingivalis* in BHI-H media were prepared and cells harvested by centrifugation. A sample of the culture fluid was taken to check for release of His-tagged CTD. Cell pellets were treated by osmotic shock to release periplasmic proteins. Following periplasmic extraction, the samples were centrifuged and a sample of the pellet taken for SDS-PAGE analysis. The cleared periplasmic fraction was then processed by Ni-NTA chromatography to capture only His-tagged proteins. As can be seen in Figure 5.4.2, no CTD protein, for any of the three constructs, could be purified following Ni-NTA chromatography of periplasmic extracts. In addition, no CTD protein was present in the culture fluid suggesting no secretion by the T9SS was occurring.

5.4.3 Whole Cell Extraction of His-tagged CTD

To determine if any expression could be detected for His-tagged CTD constructs a purification was performed using whole-cell extracts rather than using periplasmic

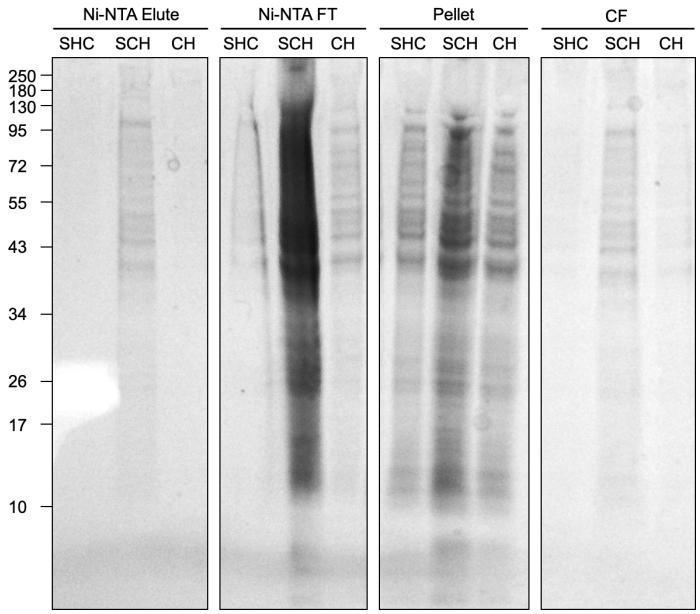


Figure 5.4.2: SDS-PAGE of Ni-NTA chromatography for periplasmic extraction of pTIO-AI constructs. samples of culture fluid, insoluble cell material, and the flow-through and elution of Ni-NTA chromatography were separated for the three pTIO-AI constructs. The expected band for RgpB-CTD, at ~10 kDa, could not be observed in any sample. Legend: FT - Flow-through, CF - Culture fluid.

fractions only. *P. gingivalis* cells carrying the correct vector were grown in 10 mL BHI-H media. Cells were harvested by centrifugation and resuspended in 1 mL PBS. The cell suspension was then lysed by sonication and again centrifuged to remove unbroken material. The cleared lysate was then loaded onto Ni-NTA agarose resin for capture of the His-tagged protein. The batch method of Ni-NTA was used with 50 µl of resin loaded into a 1.5 mL micro-centrifuge tube. Three 15 minutes washing steps were used, each with 5× resin bed volume of wash buffer supplemented with 5 mM TLCK. Elution was achieved by adding SDS-PAGE loading buffer directly to the resin followed by heating to 95 °C for 5 minutes. Samples of the Ni-NTA chromatography flow-through, wash and elution were taken for SDS-PAGE analysis.

Figure 5.4.3 shows the flow-through, wash and elution of the Ni-NTA chromatography for the SCH, SHC and CH constructs. As can be seen sample elution fraction, the above method was not successful in purifying any His-tagged material from *P.*

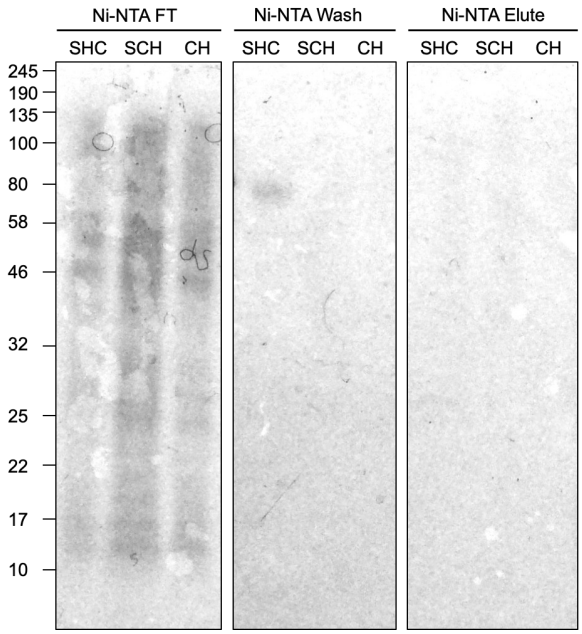


Figure 5.4.3: SDS-PAGE of Ni-NTA chromatography for whole-cell extraction of pTIO-AI constructs. Ni-NTA chromatography flow-through, wash and elution samples for each construct are shown. No bands could be observed for the RgpB-CTD at the expected molecular weight of ~10 kDa. Legend: FT - flow-through.

gingivalis lysate. To test for misfolding of the CTD constructs, the same purification was repeated with 8 M urea and 1% SDS added to buffers during lysis and purification. However, it was still not possible to identify any CTD protein using this technique.

5.5 Co-Immunoprecipitation

Whilst working on the pTIO-AI constructs, polyclonal rabbit immunoglobulins were raised by Generon against recombinantly expressed and purified RgpB-CTD. Access to this antibody provided an effective means of identifying the presence of CTD protein in a given sample, but also allowed for a co-immunoprecipitation approach to be pursued for the identification of protein-protein interactions. Co-immunoprecipitation allows for the indirect capture of interaction partners that are bound to the bait protein recognised by the antibody. These complexes can then be separated by SDS-

PAGE and potential interaction partners identified by mass spectrometry. The main advantage of using this system, over the traditional pull-down methods described above, is that complexes can be captured *in vivo*. This means the CTD proteins would pass through the native intact secretion system, which may be necessary for the capture of interaction complexes.

All co-immunoprecipitation reactions were performed using the BioVision Immunoprecipitation Kit with either the non-denaturing or denaturing lysis buffer. The initial screens were carried out using the *P. gingivalis* W50 wild-type strain lysed with non-denaturing lysis buffer. The lysate was passed over protein A/G beads alone and beads bound to Anti-RgpB-CTD. Protein A/G beads bound to Anti-RgpB-CTD with no loading of *P. gingivalis* lysate was used as a negative control. Suspensions were incubated at 4 °C for 1 hour with gentle mixing. The samples were then washed three times using 1 mL of kit washing buffer with 5 minutes incubation at 4 °C with gentle mixing for each wash step.

As can be seen in Figure 5.5.1, the Anti-RgpB-CTD antibody runs as a major band around 46 kDa and a minor band at 25 kDa. No difference could be observed between the lysate only sample and test sample containing anti-RgpB-CTD. This suggests that protein binding was due to non-specific interaction of proteins with the protein A/G beads.

To address the issue of high background binding, more stringent washing conditions were incorporated into the method. Each washing step was increased from 5 minutes to 15 minutes, again for a total of three washes. To ensure protein complexes would not be lost due to the increase in washing stringency, cells were also treated with DSP and formaldehyde cross-linkers prior to lysis. As any protein complexes would now be trapped by chemical crosslinks it was possible to use the denaturing RIPA lysis buffer for more effective release of protein from the cells. The addition of 2×SDS-loading

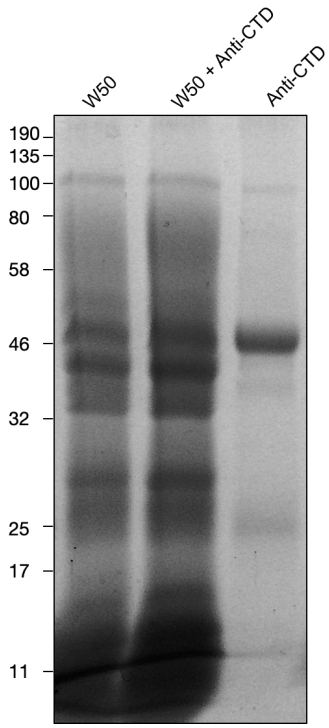


Figure 5.5.1: Co-immunoprecipitation using *P. gingivalis* W50 lysed with non-denaturing lysis buffer. W50 lysate, loaded onto Protein A/G beads without anti-CTD shows there are many non-specifically interacting proteins are captured. No additional bands can be observed in the test sample using W50 lysate captured using anti-RgpB-CTD.

buffer was used for sample elution. This was supplemented with 1×NuPAGE® Sample Reducing Agent (ThermoFisher) to ensure that the DSP cross-links would be broken allowing for the separation of individual protein-complex constituents.

The W50 derived D7 strain (ΔRgpB) was used as an alternative negative control to beads loaded with lysate only. Because the beads are still bound to anti-RgpB-CTD in this control sample it provides a more accurate picture of the non-specific interactions that occur in the absence of RgpB. Figure 5.5.2 shows the result of the co-immunoprecipitation using these conditions. It shows that no proteins were precipitated using the denaturing buffer either with or without chemical cross-linking.

As demonstrated by Glew *et al.*, [112] only a weak signal is detected for unprocessed RgpB in the W50 wild-type strain. This could be because the T9SS exports and

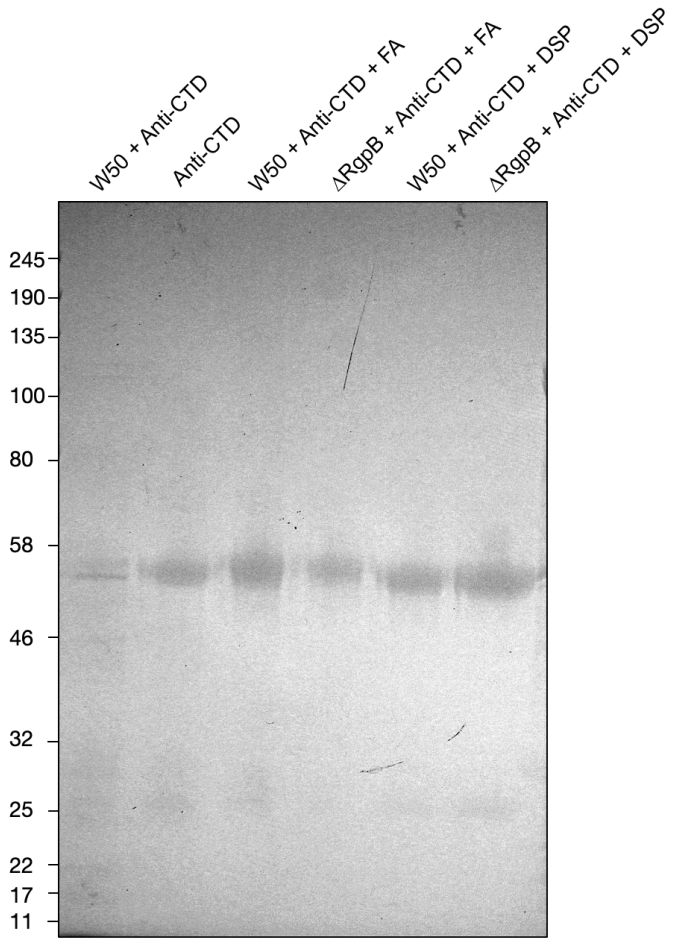


Figure 5.5.2: Co-immunoprecipitation of cross-linked *P. gingivalis* W50 WT and ΔRgpB strains. Cells were treated with DSP or formaldehyde and lysed using RIPA buffer. No bands, other than for the anti-RgpB-CTD itself can be observed.

processes cargo proteins before there is a chance of protein accumulation within the periplasm. To address this, the co-immunoprecipitation experiments were adapted to use the W50 derived ΔPorV knock-out strain. It was hoped that this would cause cargo proteins to become trapped within the T9SS and accumulate, allowing easier capture with anti-RgpB-CTD.

For these co-immunoprecipitation experiments, the ΔRgpB strain was again used as a negative control for detecting non-specific binding of protein to the anti-RgpB-CTD antibody. In addition, a rabbit anti-mouse antibody was used as a negative

control for non-specific protein binding to the antibody Fc region and the protein A/G beads. Western blot analysis was used to distinguish which protein bands were being precipitated due to interaction with anti-RgpB-CTD and which bands are being co-precipitated due to interaction with other proteins in the sample or due to interaction with the protein A/G beads.

The reaction was performed using non-denaturing lysis to ensure interaction partners were not lost. The duration of each washing step was also optimised with 10, 15 and 20 minutes being tested. Despite changes in band intensity no changes in banding pattern could be detected by changing the duration of sample washing. The precipitate of samples washed with three 15 minute washes is shown here.

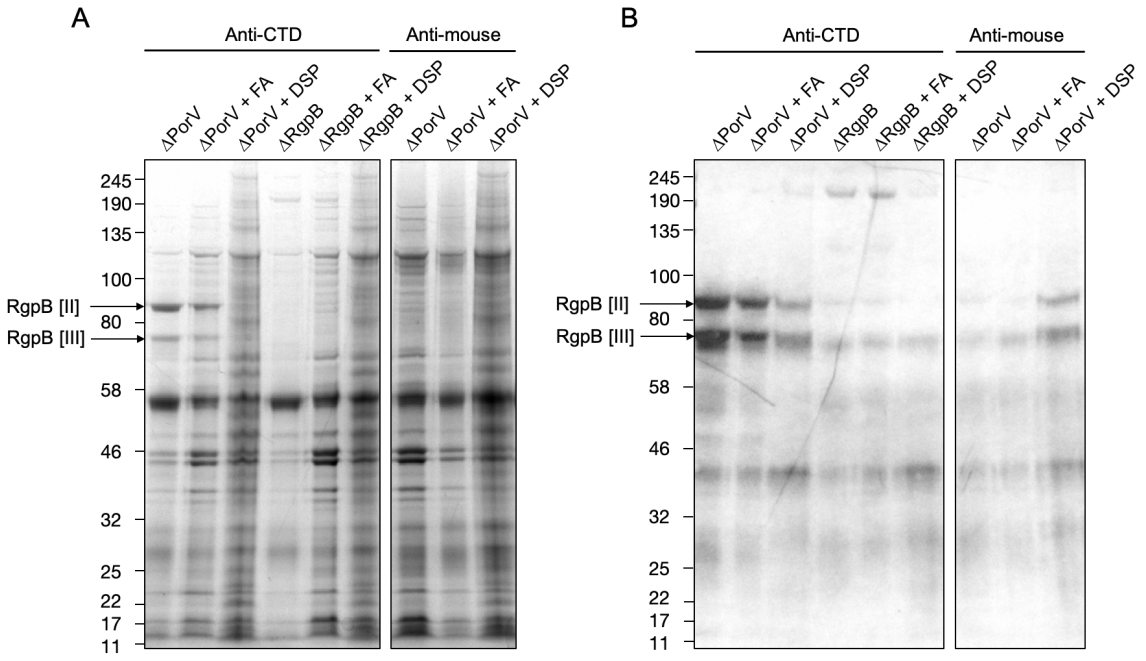


Figure 5.5.3: Co-immunoprecipitation for *P. gingivalis* ΔPorV and ΔRgpB strains with non-denaturing lysis conditions. (A) SDS-PAGE analysis of co-immunoprecipitation elution and (B) corresponding western blot analysis of the samples detected with anti-RgpB-CTD or anti-mouse. Using the ΔPorV mutant it was possible to capture the [II] and [III] forms of pro-RgpB, previously identified by Glew *et al.,* [82]. Anti-mouse and ΔRgpB were used as a negative control samples.

As can be seen in Figure 5.5.3 it was possible by using the ΔPorV mutant, to

precipitate the [II] and [III] proteolytically processed pro-RgpB, previously identified by Glew *et al.,* [82] using ΔPorU and ΔPorV mutants. Significantly more protein bands were found to co-precipitate following chemical cross-linking with formaldehyde or DSP. However, comparison of these bands to the ΔRgpB and anti-mouse negative controls reveals that the co-precipitated bands in the ΔPorV are not due to non-specific interactions.

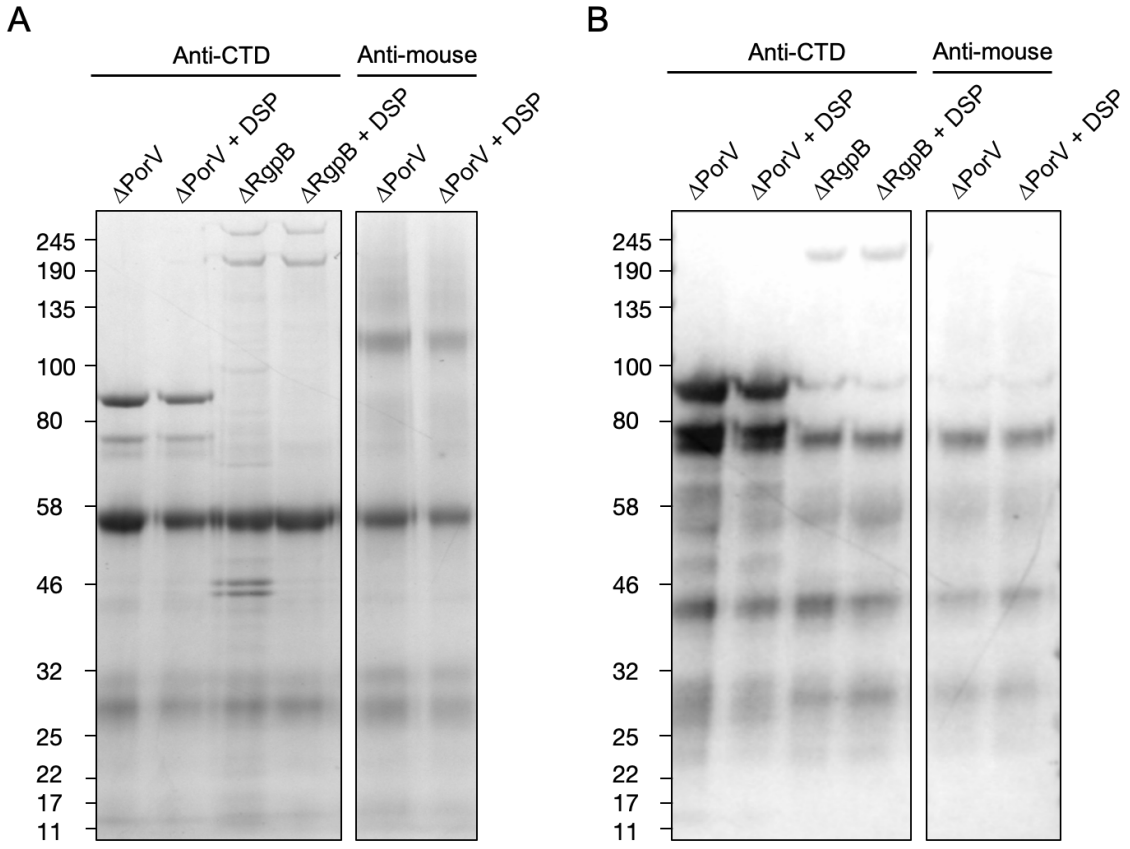


Figure 5.5.4: Co-immunoprecipitation of *P. gingivalis* ΔPorV and ΔRgpB strains with denaturing lysis conditions. (A) SDS-PAGE analysis of co-immunoprecipitation elution and (B) corresponding western blot analysis of the samples detected with anti-RgpB-CTD or anti-mouse. Anti-mouse and ΔRgpB were used as a negative control samples.

Because it was possible that the high number of non-specific bands, particularly in the DSP cross-linked sample, were causing genuine interactions to be masked the reaction was repeated using denaturing lysis buffer. Given that the banding pattern

was very similar for the uncross-linked and formaldehyde cross-linked samples (Figure 5.5.3), the formaldehyde treatment was not repeated. The optimised washing conditions of three 15 minute washes was used. Figure 5.5.4 shows that Pro-RgpB was still precipitated under denaturing conditions. There was no detection of co-precipitating bands due to interaction with the Pro-RgpB protein.

5.6 Discussion

The experimental techniques described in this chapter were performed with the aim of identifying components of the T9SS that are able to interact with the CTD of T9SS cargo proteins. Identification of CTD interaction partners in *P. gingivalis* would not only advance understanding of the functional role of proteins comprising the T9SS, but would also allow for such interactions to be characterised in greater detail. Following the identification of novel interactions, structural studies for the complex were intended to gain a better understanding of the structural motif present in the cargo CTD that mediates transport through the system.

When this study was initiated no interaction between cargo CTDs and components of the T9SS had been characterised. It still remains unclear what components of the T9SS are capable of directly interacting with the CTD, if there are any periplasmic chaperones and in what order the cargo is transferred between components of the system. To identify these potential interactions a range of experimental protein interaction screening techniques were applied. Techniques were selected that screen for interactions in a non-specific way to allow identification of novel interaction and isolation of unknown interaction partners not currently considered part of the T9SS machinery. Therefore, methods that test specific interactions between two selected proteins, such as a bacterial-2-hybrid system, were not employed here. The pull-down and co-immunoprecipitation techniques were selected as they allow a large pool of

potential interaction partners to be tested simultaneously with the ability to capture transient interactions by the addition of cross-linking treatment. These techniques, as described here, were employed unsuccessfully towards this aim. In light of more recent publications, describing the structure and function of the T9SS, many of the limitations in the methodology can be reasoned.

5.6.1 Limitations of Ni-NTA Pull-Down Methodology

Ni-NTA pull-down experiments were designed to use recombinant His-RgpB-CTD mixed with *P. gingivalis* lysate. This method relies on solubilised protein from *P. gingivalis* being exposed to binding sites on immobilised CTD protein. The lysis method employed here involved sonication of cells suspended in PBS containing triton X-100 to allow the solubilisation of membrane proteins. It is possible that the lysis method used causes dissociation of large protein complexes comprising the T9SS. For example, the purification of a 50 nm diameter protein ring, comprised of PorK and PorN was achieved using osmotic shock, and the addition of 1% DDM for membrane solubilisation, to ensure the complex remained intact [115]. It may be that intact complexes are necessary to mediate interactions with the CTD. Adjustment of the method to utilise chemical lysis may ensure protein macrostructure is maintained.

The method of protein extraction may also be ineffective at releasing membrane protein components of the system. For example, glew *et al.*, [112] were able to separate complexes of PorV bound to the cargo proteins HagA, Kgp/RgpA, RgpB, PG2102 and PG2172 in a PorU deletion mutant. The protein was extracted using BN-PAGE lysis buffer containing 1% DDM and separated by 2D BN-PAGE. In addition, extraction of PorV from *P. gingivalis* using 1% LDAO as a solubilising agent has been described [107]. However, the protein extraction method described here utilised triton X-100 for membrane protein solubilisation. This agent may be ineffective at solubilising PorV

and other membrane protein components of the T9SS.

Transient interactions are often not observed in pull-down experiments due to the chemical stress of sample washing. Given that the cargo protein may need to be transferred between multiple binding partners it is possible that transient interactions occur during the secretion process. In an attempt to capture potential protein interactions that are being lost during sample washing a cross-linking step was incorporated into the pull-down experiment. An additional advantage of this technique is that denaturing conditions can be following cross-linking treatment to solubilise all proteins in the sample. However, it is possible that cross-linking treatment causes the formation of large insoluble complexes that are lost during Ni-NTA loading and washing steps.

5.6.2 Limitations of pTIO-AI Expression System

The pTIO-I plasmid was developed as a vector suitable for general use, easy handling and high electrotransformation efficiency to facilitate molecular biology studies in *P. gingivalis* [230]. Previously available plasmids for use in *P. gingivalis* suffered from low electrotransformation efficiency, vector instability and large vector size. However, the pTIO-I plasmid was not developed for recombinant expression of proteins in *P. gingivalis*. It was further developed by Dr Mark Roberts to include an RgpA promoter to allow for recombinant expression of genes within *P. gingivalis*. The success of this expression system was confirmed by the insertion of the β-galactosidase gene and subsequent analysis of β-galactosidase activity within cells. The pTIO-AI plasmid was gifted by Dr Mark Roberts for use in this study.

Given that this vector has been shown to remain stable within *P. gingivalis*, is easy to manipulate and now carries a promoter allowing expression of a given gene of interest, it was selected as an ideal vector for use in this study. However, despite successful assembly and electrotransformation into *P. gingivalis*, no expression of the

CTD constructs could be detected. This suggests that the constructs were either not expressed, they were quickly degraded or that the expression level was too low to be detected by SDS-PAGE. Detection of β-galactosidase expression was achieved by checking cellular β-galactosidase activity which is more sensitive that SDS-PAGE analysis. This supports the theory that expression of the CTD constructs is too low to be detected by SDS-PAGE. By scaling-up the volume of *P. gingivalis* culture loaded onto Ni-NTA resin it may be possible to enhance the concentration of His-tagged CTD protein. More extensive testing of promoter efficiency and testing of alternative promoters may be necessary before this expression can be used for effective recombinant protein expression in *P. gingivalis*.

At this time anti-RgpB-CTD became available making co-immunoprecipitation possible. Given that co-immunoprecipitation also facilitates *in vivo* capture of protein complexes, but without the need for optimisation of protein expression, this methodology was not developed further. Further justification of this decision is that it the same issues arising during the co-immunoprecipitation experiments would likely occur as the CTD is exposed to the system in a very similar way. However, the potential issue of competitive binding between components of the secretion system and the anti-RgpB-CTD antibody would have been circumvented. This spatial separation of the His-tag epitope, present on the pTIO-AI expressed protein, from the CTD may have allowed more efficient capture of protein by Ni-NTA chromatography relative to Co-immunoprecipitation of native protein.

5.6.3 Limitations of the Co-Immunoprecipitation Methodology

Co-immunoprecipitation experiments were adapted for use in a ΔPorV strain of *P. gingivalis* W50 in the hope that this would interrupt the secretion process causing

cargo proteins to become blocked within the secretion system. Ishiguro *et al.*, [107] describe the accumulation of Arg-gingipains within the periplasm in a PorV deletion mutant derived from *P. gingivalis* W83 suggesting that secretion is disrupted in the absence of PorV. Therefore, co-immunoprecipitation experiments were adapted for use in a ΔPorV strain of *P. gingivalis* W50 in the hope that this would interrupt the secretion process. As shown in Figure 5.5.3 it was possible to capture pro-RgpB by immunoprecipitation from this strain. However, no co-precipitating proteins could be identified.

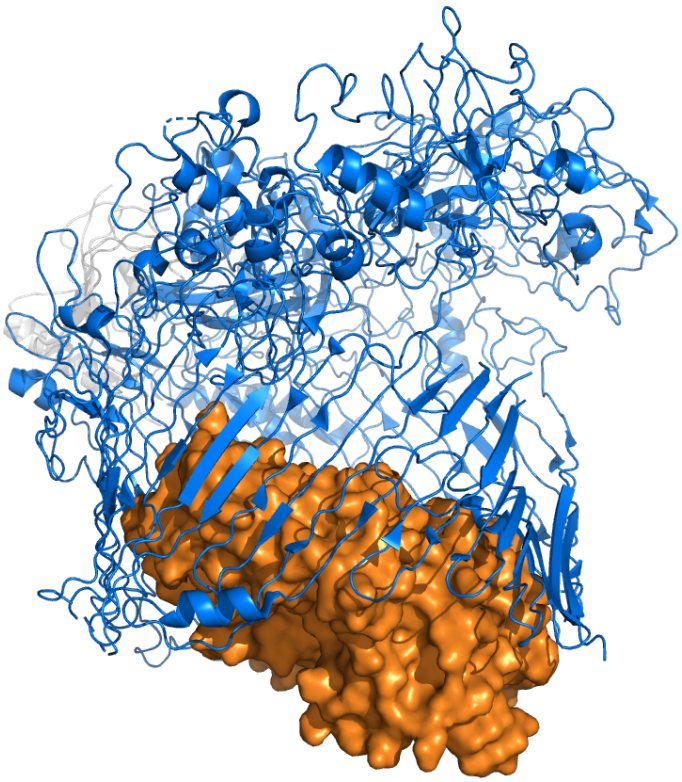


Figure 5.6.1: In the absence of PorV the SprA pore is blocked by a plug protein. Structural model of SprA (blue) in complex with the T9SS plug protein (orange). In the presence of the plug protein prevents the export of proteins across the outer membrane.

More recently a cryo-EM structure for SprA, the Sov homolog in *Flavobacterium johnsoniae*, has been determined [113]. The protein forms a large transmembrane β-

barrel with a cap at the extracellular end. A lateral opening is present which creates a cavity between the periplasm and the outer membrane. The structure is presented in two states, the first has PorV bound to the lateral opening with three loops exposed to the periplasmic cavity. The second state contains a plug protein bound to periplasmic opening of SprA. The evidence presented by the authors suggest that in the absence of PorV the porin of the T9SS is maintained in a closed state as shown in Figure 5.6.1. This would prevent the capture of any Sov-RgpB complexes by co-immunoprecipitation in the ΔPorV mutant. This mechanism of inactivation may also interrupt the interaction of other system components with RgpB and explain why protein complexes could not be captured. It is possible that the use of other mutants in co-immunoprecipitation experiments, such as porU or PorZ deletion mutants, in which cargo proteins appear to accumulate in a PorV bound state at the outer membrane, would cause cargo proteins to be blocked within an open system. This may have allowed protein complexes involving pro-RgpB to be captured by co-immunoprecipitation.

5.6.4 The Role of PorV in Cargo Protein Transport

To date, there is only one study demonstrating an interaction between the T9SS system and T9SS cargo proteins [112]. The authors were able to demonstrate using 2D BN-PAGE followed by LC-MS, in a ΔPorU strain, that PorV interacts with cargo proteins of the T9SS. They were able to isolate PorV in complex with HagA, Kgp/RgpA, PG2102 and PG2172. Although they did not directly demonstrate that this interaction was being mediated by the CTD, all proteins isolated were in their unprocessed state possessing both pro-peptides and CTDs. That the authors required a PorU deletion strain to isolate interaction partners is consistent with the hypothesis that processing in wild-type cells is too fast to allow capture of complexes containing unprocessed

protein.

The authors also isolated complexes from a strain possessing a deletion mutation in the PorU cleavage site in the RgpB gene. It was observed in this mutant that unprocessed RgpB would accumulate in a PorV bound state on the surface of the cell. This again further suggests that the interaction between RgpB and PorV is being mediated by the CTD. It also suggests that the use of a PorU deletion mutant for co-immunoprecipitation experiments would have allowed capture of the PorV interaction.

5.6.5 Selection of CTD Interaction Partners

Although it is speculated that the interaction between PorV and RgpB is mediated by the CTD, given that it is the only component common to all cargo proteins, it has not been directly demonstrated. The ambition of this thesis is to directly demonstrate an interaction between the CTD of RgpB and the T9SS. Therefore, PorV was selected as the candidate interaction partner suitable for further analysis. To allow for a thorough examination of this interaction no further experiments were conducted to identify other components of the T9SS that may interact with the RgpB-CTD protein.

5.7 Summary

This chapter describes the application and optimisation of non-specific protein interaction screening using RgpB-CTD tested against proteins from *P. gingivalis*. The aim of this was to isolate candidate proteins, one of which would be taken forward for further characterisation and structural studies. The initial approach involved the addition of His-tagged RgpB-CTD to *P. gingivalis* lysate followed by capture using Ni-NTA chromatography. This was adapted to incorporate chemical cross-linking in the hope of trapping transient interactions. This approach was unsuccessful in isolat-

ing interaction partners. Therefore, an *in vivo* approach was developed using pTIO-AI for expression of tagged RgpB-CTD in *P. gingivalis.* However, no protein could be identified by SDS-PAGE using this system and it requires further optimisation. Co-immunoprecipitation using anti-RgpB-CTD was then applied to capture native RgpB passing through the T9SS. Despite using chemical cross-linking no RgpB could be captured in wild-type *P. gingivalis.* The technique was adapted to use ΔPorV with the aim of reducing export efficiency and causing cargo proteins to become trapped within the T9SS. However, Lauber *et al.*, [113] demonstrated that when PorV is deleted a plug protein blocks the T9SS secretin and may prevent CTD proteins interacting with the T9SS. Given that no viable interaction partners were identified using these techniques, PorV was selected for further characterisation based on evidence from 2D-BNPAGE and LC-MS that PorV interacts with cargo proteins from the T9SS.

Chapter 6

Characterisation of RgpB-CTD Interaction With PorV

6.1 Chapter Aims

The aim of this chapter was to characterise the hypothesised interaction between RgpB-CTD (described in Chapter 4) and PorV (selected in Chapter 5) using *in vitro* methods for studying protein-protein interactions (PPI). This would be the first *in vitro* characterisation of an interaction between the CTD of any T9SS cargo protein and a component of the T9SS. This information would further our understanding of how cargo proteins are recognised and processed by the T9SS. Furthermore, if a structural model of the complex could be developed this may provide new possibilities for the development of therapeutic molecules that prevent CTD recognition by the T9SS.

6.2 Rational Design of Recombinant PorV

To achieve the chapter aims it was necessary to develop a method for the expression and purification of PorV protein to be used for *in vitro* PPI experiments. The objective was to generate purified PorV sample that could be used in pull-down experiments with purified RgpB-CTD to directly demonstrate the interaction between the two proteins. Point mutations would then be generated for each protein to generate low resolution restraints that could be used to define an interaction interface between the two proteins. The *P. gingivalis* W50 PorV (PG0027) was selected as the RgpB-CTD construct described in this thesis was also taken from *P. gingivalis* W50.

6.2.1 Export Prediction

Selection of the correct protein sequence is an important first step in the design of protein constructs for recombinant expression. Analysis of the protein sequence provides insight into the most appropriate system for expression and purification. It has previously been demonstrated that PorV is localised to the outer membrane and is predicted to possess a β-barrel fold [108]. Bacterial outer membrane β-barrel proteins are known to possess N-terminal signal peptides for transfer from the cytoplasm to the periplasm via the Sec system [232]. Therefore, the primary sequence of *P. gingivalis* W50 PorV was uploaded to the SignalP-5.0 server [159] for the prediction of signal peptides and cleavage sites. As shown in Figure 6.2.1 PorV is predicted to contain an N-terminal signal peptide (residues 1-27) for Sec transport with peptide cleavage occurring between residues A^{27} and Q^{28}. For expression of PorV in *E. coli,* residues 28-391 were taken for assembly into pASK-IBA2 or modified pET-28, carrying the *E. coli* OmpA Sec signalling peptide. The OmpA signalling peptide was selected as it provides a highly effective signal that has been used extensively for recombinant expression of periplasmic and membrane proteins in *E. coli* [233].

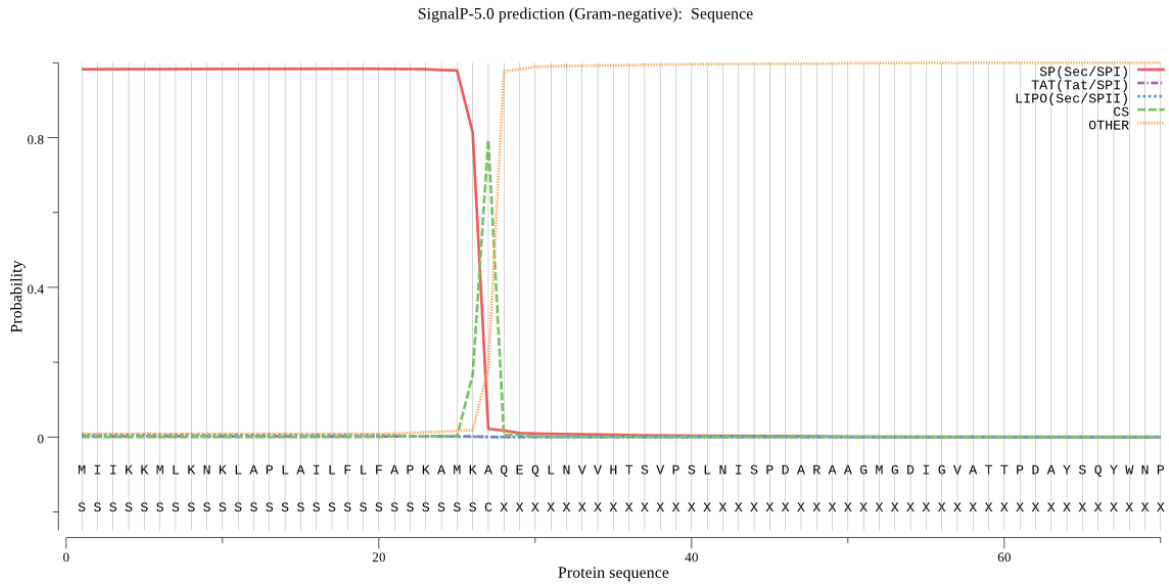


Figure 6.2.1: SignalP server result for porV protein sequence. This predicts that PorV contains a Sec signal peptide spanning residues 1-27. Peptide cleavage is predicted to occur between residues A27 and Q28.

6.2.2 Tertiary Structure Prediction

When work for this chapter began no structural model was available for the PorV protein. Therefore, a prediction of the protein tertiary structure was calculated to determine the location of N- and C-terminal residues. This was used to ensure the addition of purification tags would not block the putative interaction site, expected to be localised to the outer face of PorV. Tertiary structure modelling was performed using the online Phyre2 server [168] which predicts protein folds via sequence alignment to known protein structures. As shown in Figure 6.2.2, PorV is predicted to contain 14 transmembrane β-strands with an N-terminal coil passing through the centre to the extracellular face. It was predicted that the C-terminus is located at the periplasmic side of the outer membrane. This orientation suggests that a C-terminal purification tag should not interfere with CTD interactions. The length of the N-terminal plug also suggests that the addition of a purification tag to the N-terminus may also remain distal from putative interaction sites in the inter-strand loops of PorV. Therefore,

constructs of PorV were assembled using both N-terminal and C-terminal tags.

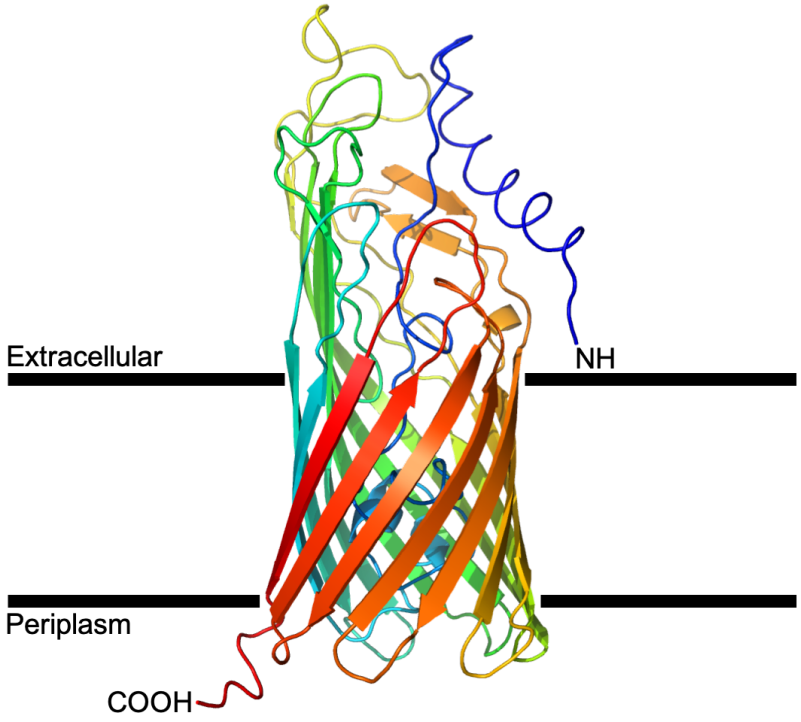


Figure 6.2.2: Phyre2 3D structural model prediction of *P. gingivalis* W50 PorV. The predicted model calculation suggests PorV is composed of 14 transmembrane β-strands with an N-terminus coil forming a plug passing to the extracellular surface. The C-terminus be exposed at the periplasmic face.

6.2.3 Construct Design

PorV$^{28\text{-}391}$ was cloned into pASK-IBA2C. This vector was selected as it contains the OmpA Sec signalling peptide for transport of fusion protein across the inner-membrane, a C-terminal Strep-tag®II purification peptide allowing for highly efficient purification of functional protein [234] and inducible tetracycline promoter/operator providing highly stringent control of protein expression [235].

Additional PorV constructs were made with the assistance of Sunjun Wang (Garnett lab, KCL) using pET28(b) modified to contain the OmpA Sec signalling peptide; renamed as pET28(b)-OSS. This vector was selected as it allows for purification of

function protein using Ni-NTA chromatography and IPTG controlled expression using a T7 promoter. Via the addition of the OmpA signalling peptide this vector could be used to produce gene products that translocate to the outer membrane. PorV$^{28\text{-}391}$ was ligated into pET28(b)-OSS carrying either an N-terminal or C-terminal 6XHis-tag. Both fusions were created to circumvent potential issues with protein folding arising from the addition of an artificial tag to either the C- or N-terminus. A schematic highlighting the orientation of gene fragments for each of the three constructs is shown in Figure 6.2.3.

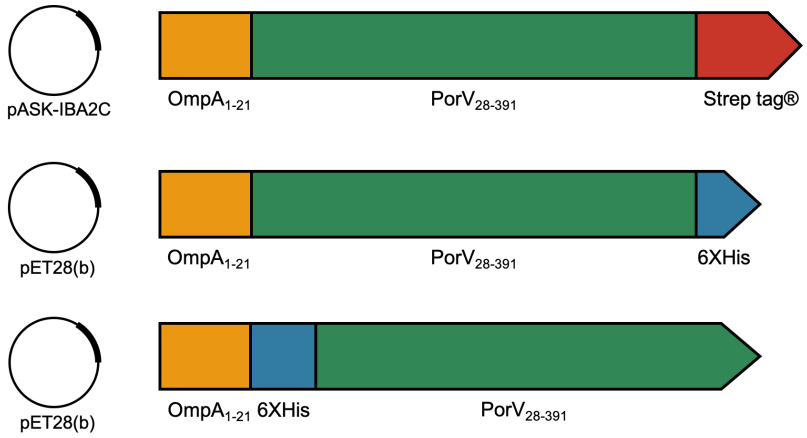


Figure 6.2.3: Design of PorV constructs for recombinant expression in *E. coli*. Vectors used for expression are shown on the left. The arrangement of gene fragments for each construct is shown on the right.

6.3 Protein Expression and Purification

6.3.1 Selection of Cell Type for Protein Expression

Proteins that localise to biological membranes are notoriously difficult to express and purify in comparison to soluble globular proteins. As such, a range of conditions for growth and expression were tested for the recombinant PorV constructs to improve protein yields.

The first variable tested was the selection of the cell type used for expression. *E. coli* strains were selected as they are Gram-negative bacteria, making it appropriate for the expression of an outer membrane protein from *P. gingivalis*, which is itself Gram-negative. In addition, *E. coli* cells are easy to handle and economically viable at the large-scale required to yield membrane protein in adequate quantities.

BL21(DE3) cells were tested as this strain is effective at expressing proteins carried by pET vectors under the control of the T7 promoter and are compatible with the tetracycline promoter/operator system [212]. The BL21(DE3) derived Lemo21(DE3) strain was also tested as this strain allows more stringent control of the T7 promoter [236]. This strain carries pLemo, containing the *lysY* gene, which encodes lysozyme - the natural inhibitor T7 RNA polymerase. Expression of lysozyme is controlled via the addition of L-rhamnose to the growth media allowing finer control of T7 system expression. This is effective at reducing the formation of inclusion bodies or reductions in cell growth due to protein toxicity; both of which are issues common to membrane protein expression.

The final cell type that was tested is also derived from BL21(DE3). The C41(DE3) strain contains at least one mutation that prevents cell death from the expression of many toxic proteins [237]. Therefore, this strain was selected for the same reasons as for the Lemo21(DE3) strain. However, unlike Lemo21(DE3), expression is not tuneable when using C41(DE3). Despite this many publications describing the purification of bacterial outer membrane proteins found C41(DE3) to be a successful and effective host [237]. Both the Lemo21(DE3) and C41(DE3) strains possess the same host features and benefits of the BL21(DE3) parent strain.

6.3.1.1 pASK-IBA2C Cell Type Screen

pASK-IBA2C was tested first as the tetracycline promoter/operator system and Strep-tag II is a highly effective system for expression of membrane proteins. Expression was tested using C41 and BL21 cells, but not Lemo21(DE3), as this strain is incompatible with the tetracycline promoter system. Cultures of C41 and BL21 carrying the pASK-PorV plasmid were grown to an OD_{600} of ~0.4 and protein expression induced via the addition of 0, 1 or 2 µg ml^{-1} of anhydrotetracycline. This tetracycline derivative is capable of activating the Tet promoter on the pASK-IBA2C plasmid but does not exhibit any antibiotic activity. This allows the promoter to be activated without slowing bacterial growth and protein expression. Following induction the cultures were grown at 18 °C for 16 hours and a 1 ml sample taken from each culture and lysed using BugBuster® Protein Extraction Reagent (ThermoFisher Scientific). Following lysis half of the sample was centrifuged to clear insoluble material. Whole cell lysate and soluble fraction lysate was separated by SDS-PAGE and immunoblotted using Strep-Tactin®-HRP (IBA-Lifesciences) for the detection of Strep-tag®II fusion proteins. As shown in Figure 6.3.1A only one band at the 25 kDa mark could be detected, despite an expected molecular weight of ~46 kDa for the PorV construct, and that this band was also present in uninduced cells. To determine if this band could correspond to leaky expression of PorV, or a fragment of the full protein, a lysate fraction was analysed from BL21(DE3) carrying the empty pASK-IBA2C vector (Figure 6.3.1B). This showed that the same band is present in absence of the fusion protein confirming that is not the PorV protein.

6.3.1.2 pET28 Cell Type Screen

For cell type screening with pET-28-OSS-PorV-His and pET-28-OSS-His-PorV was performed using a similar method as described above. Lemo21 (DE3) were also used to

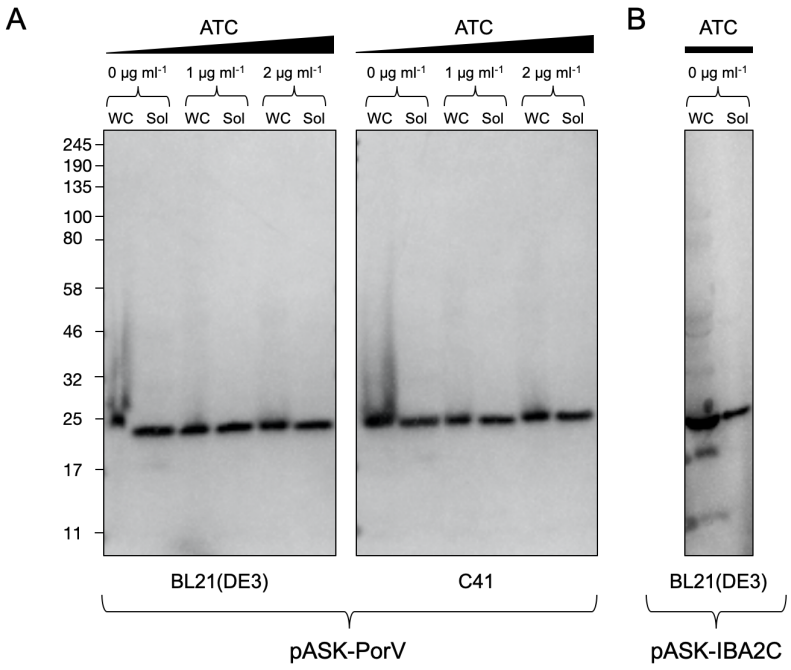


Figure 6.3.1: pASK-PorV expression trial. (A) Comparison of cell types for the expression of PorV-StrepII fusion protein expressed in pASK-IBA2C. 0, 1 and 2 µg ml^{-1} of anhydrotetracycline was used to induce protein expression. (B) Analysis of expression background in BL21(DE3) carrying the empty pASK-IBA2C vector. Legend: ATC - anhydrotetracycline, WC - whole cell lysate, Sol - Soluble cell lysate.

allow more stringent control of IPTG induction. For the Lemo21(DE3) cultures 0, 100, 250, 500, 750, 1000 or 2000 µM L-Rhamnose was added. Increasing concentrations of L-Rhamnose reduces the total level of protein expression but should reduce the tendency for inclusion body formation. For all cultures 0.4 mM IPTG was used to induce cells at an OD_{600} of ~0.4. Following 16 hour incubation at 18 °C cells were then treated using the same method as for the pASK-PorV samples. The samples were immunoblotted using anti-His-HRP (ThermoFisher Scientific) for the detection of His-tag fusion proteins.

For the C-terminal tagged construct no bands could be detected in the western blot in any sample (Figure not shown). For the N-terminal tagged construct one major band just below the 46 kDa marker could be detected (Figure 6.3.2). This band was

most prominent in the BL21(DE3) sample and Lemo21(DE3) with 0 or 100 µM L-Rhamnose. Given the expected molecular weight of ~44 kDa for the PorV construct this band was taken to correspond to the PorV construct. Based on the observed levels of expression pET-28-OSS-His-PorV and BL21(DE3) was taken forward as the best construct and cell type, respectively, for expression of the PorV construct.

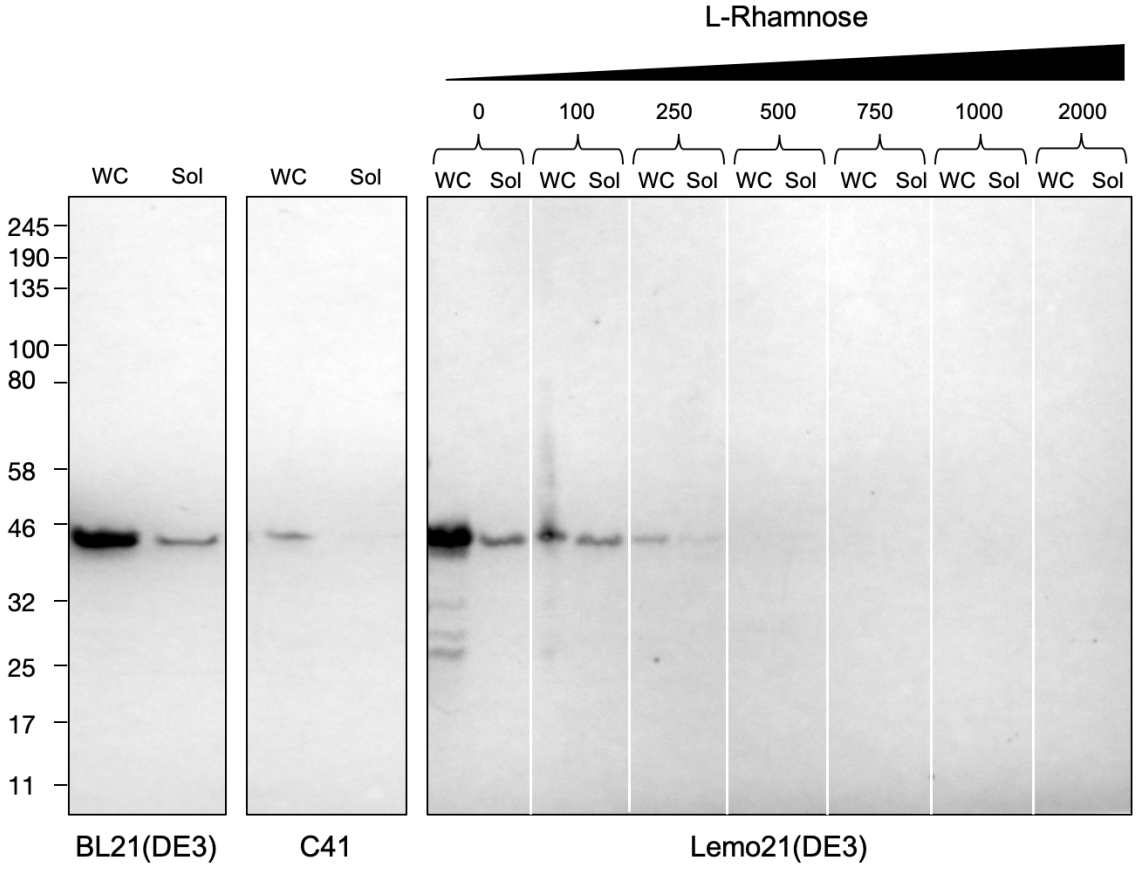


Figure 6.3.2: Comparison of cell types for expression of pET28-OSS-His-PorV. pET28-OSS-His-PorV was expressed in BL21(DE3), C41 and Lemo21(DE3) cells using 0.4 mM IPTG for protein induction. L-Rhamnose optimisation was also used for optimisation of Lemo21(DE3) expression. Legend: WC - whole cell lysate, Sol - Soluble cell lysate.

6.3.2 Optimisation of Expression Parameters

Preliminary screening of PorV expression showed that the protein was prone to the formation of insoluble inclusion bodies. Therefore, extensive screening of expression conditions was performed to identify conditions conducive to the expression of correctly folded protein.

Alterations in growth media were analysed as this can have a significant impact on protein expression levels without large increases in preparation costs [238]. Rich media (LB and 2XYT) was compared using a range of culture densities prior to IPTG induction. OD_{600} values of 0.4, 0.8 and 1.2 were used for cells grown in LB. Given that 2XYT can support higher cell densities OD_{600} values of 0.8, 1.2 and 1.6 were used. Cells grown in M9 minimal media will grow slower than cells grown in rich media. Although this may reduce the total level of protein expression it was hoped that the slower growth may help to prevent inclusion body formation. Following induction cells were grown at 18 °C for 16 hours. The same method as before was used for sample lysis and separation. Immunoblots were performed using anti-His-HRP. Only soluble lysate fractions were analysed.

As shown in Figure 6.3.3A expression could be observed for all OD_{600} values using LB and 2XYT. However, no expression could be observed when using M9 minimal media. The 2XYT sample, induced at OD_{600}~ 0.8, was taken to be the most effective expression condition of those tested. This set of parameters was taken forward into IPTG concentration screening. Cultures were grown in 2XYT to an OD_{600}~ 0.8 and induced with 0, 50, 100, 200, 400 or 600 µM IPTG. As shown in 6.3.3B the most effective concentration range for IPTG is between 50-200 µM. Therefore, 50 µM was selected as the concentration of IPTG to use for large-scale expression.

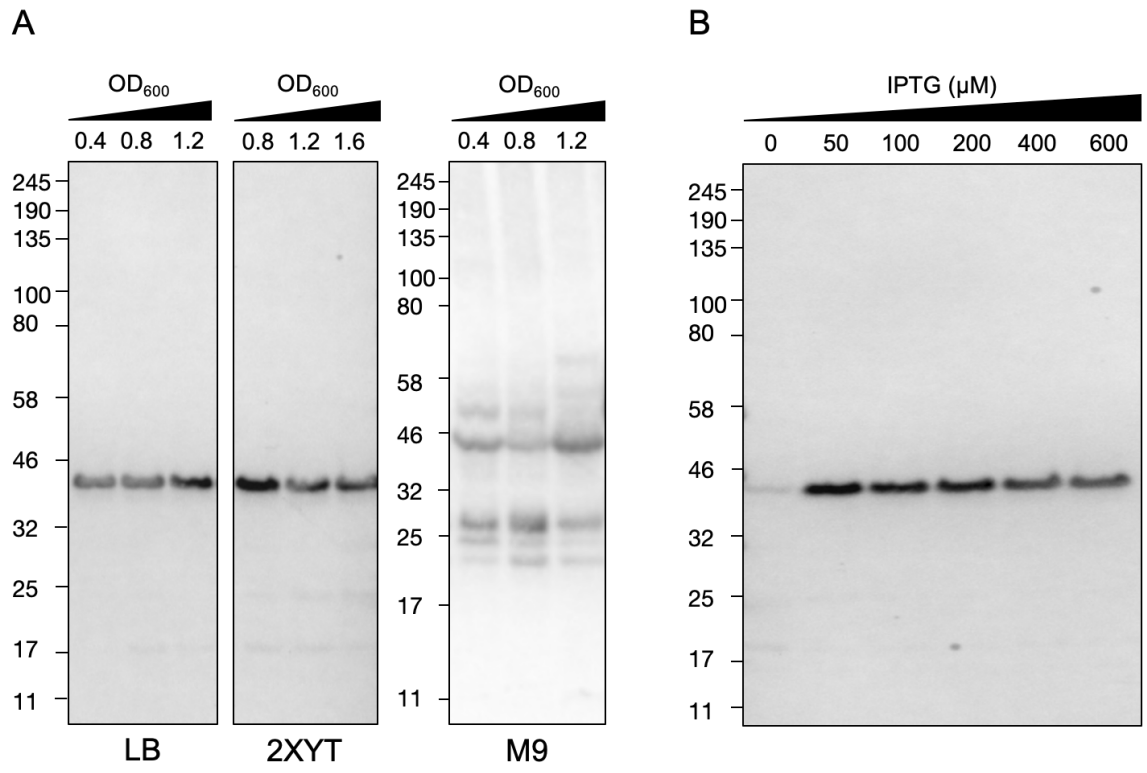


Figure 6.3.3: Expression parameter screening for pET28-OSS-His-PorV in BL21(DE3). (A) Screening of growth media composition and induction OD_{600}. (B) Screening of IPTG concentration for the best media and induction OD_{600} parameters (2XYT and OD_{600} ~0.8).

6.3.3 Detergent Screening

For the membrane proteins to be extracted for the native biological membrane detergents that mimic the lipid bilayer are typically used to disrupt the membrane and surround the hydrophobic transmembrane portion of the protein. The exact detergent used requires careful selection as they can also cause denaturation of the membrane protein or fail to solubilise the protein at all. Furthermore, some detergents cause the solubilised protein to lose its functionality.

The majority of detergents used for the solubilisation of membrane proteins are non-ionic [239] as ionic detergents, such as SDS, are harsher to the membrane and protein. For the solubilisation of PorV the non-ionic detergents OG, LDAO, C8E4 and

DDM were used. In addition, solubilisation using styrene–maleic acid (SMA) was also tested as proteins solubilised via this method are retained in lipid discs which is more similar to their native environment [169]. This has been shown to increase thermostability and protein functionality relative to detergent solubilised proteins [240–242]. A range of different SMA co-polymers are available but SMA2000 has been shown to be the most effective for protein solubilisation and is the most commonly used co-polymer [169].

Extracted membranes containing the PorV were solubilised using C8E4 (1%), LDAO (1%), DDM (2%), OG (2%) and SMA2000 (2.5%) at room temperature for 1 hour. The samples were then ultracentrifuged to clear the insoluble material. Soluble and insoluble fractions from each sample were then separated via SDS-PAGE and immunoblot analysis performed using anti-His-HRP. This was performed in triplicate and used to calculate solubilisation efficiency of each additive (Figure 6.3.4). This showed that LDAO is significantly more effective at solubilising PorV than any other additive ($P < 0.00001$, ANOVA). Therefore, LDAO was selected for solubilisation of PorV in large-scale purifications.

6.3.4 Large-Scale Protein Expression

The optimised expression parameters were used to prepare PorV protein from pET-28-OSS-His-PorV grown in BL21(DE3) cells. Cells were grown to an OD_{600} of ~0.8 in 2XYT media at 37 °C and induced with 50 µM IPTG. The cultures were then grown at 18 °C for 16 hours and the cells harvested by centrifugation. For the large-scale preparation sonication was used in place of BugBuster lysis. Membranes were collected using ultracentrifugation and solubilised using 1% LDAO. After incubation at room temperature for 1 hour the suspension was ultracentrifuged to clear insoluble material. The resulting supernatant was used for Ni-NTA chromatography.

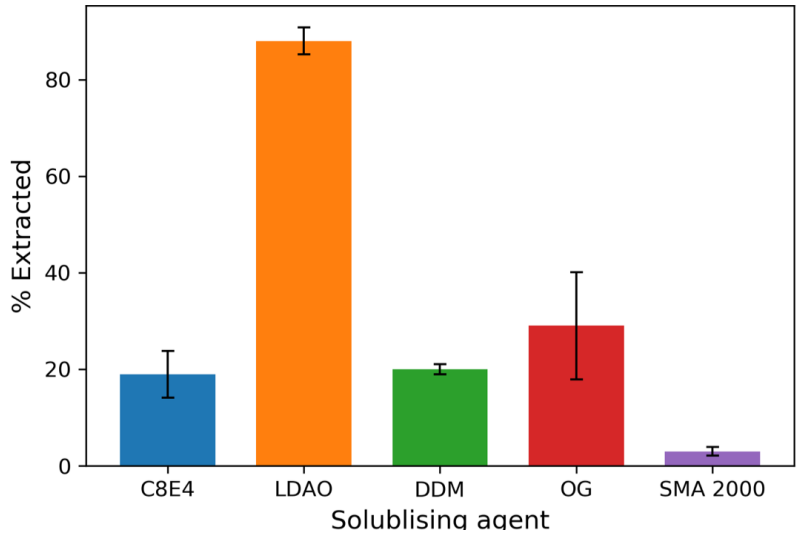


Figure 6.3.4: PorV solubilisation screening. Comparison of solubilisation efficiency for each solubilisation additive showing LDAO is the most effective solubilisation agent for PorV ($P < 0.00001$, ANOVA). Legend: S - Soluble fraction, P - Insoluble pellet.

6.3.5 Ni-NTA Chromatography

The solubilised membrane suspension was loaded onto an equilibrated poly-prep gravity flow column containing 1 ml Ni-NTA agarose resin (ThermoFisher Scientific). The was washed using 20 column volumes of Ni-NTA wash buffer and eluted with a series of 2 ml factions using Ni-NTA elution buffer. All Ni-NTA buffers were supplemented with 0.1% LDAO to maintain protein solubility. The fractions were separated by SDS-PAGE and analysed via western blot using anti-His-HRP (Figure 6.3.5). This showed that some PorV protein will dissociate from the column during washing but the majority of the protein is eluted during the first 6 ml of elution.

6.3.6 Size Exclusion Chromatography

The first 8 ml of Ni-NTA chromatography elute (E1-4) was pooled and concentrated to a volume of 0.5 ml. As shown in Figure 6.3.5A the protein purify was poor following Ni-NTA chromatography alone. Therefore the sample was further purified by size

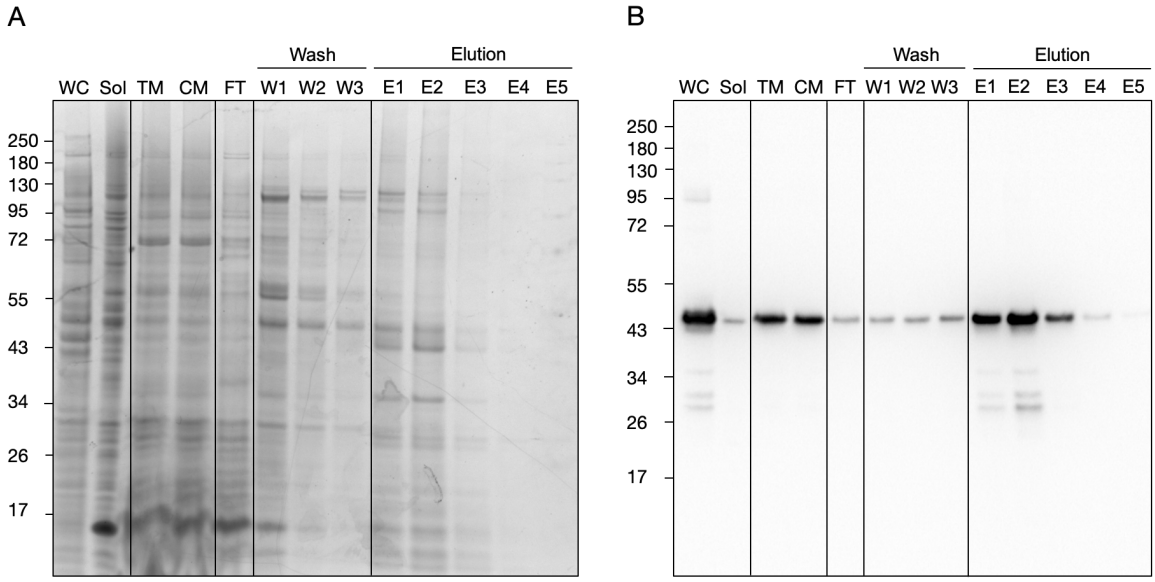


Figure 6.3.5: Ni-NTA chromatography purification of His-PorV. (A) SDS-PAGE analysis of purification samples and chromatography elution profile. (B) corresponding western blot processed with anti-His-HRP. Legend: WC - whole cell lysate, Sol - Soluble cell lysate fraction, TM - Total membrane fractions, CM - Cleared membrane fraction, FT - Flow-through during sample loading, W1-3 - 5 ml wash fractions, E1-5 - 2 ml elution fractions.

exclusion chromatography to remove contaminant proteins and imidazole. Due to the molecular expected molecular weight of a His-PorV:LDAO complex (~61 kDa) a HiLoad 16/600 Superdex 200 pg column was used as it provides good resolution for globular proteins between 10-600 kDa. Figure 6.3.6A shows that there are four major peaks that elute after the void volume of 45 ml. To determine which of these peaks corresponds to His-PorV an additional purification was performed using BL21(DE3) carrying pET28(b)-OSS with no fusion protein. Comparison of the two absorption spectra revealed that there are unique peaks at the void volume and at ~69 ml in the His-PorV purification sample. Separation of 2 ml elution fractions representing the peak at 69 ml by SDS-PAGE and anti-His-HRP western blot showed the presence of a major band just above the 43 kDa marker (Figure 6.3.6B-C). Additional smaller bands could be observed in several fractions which may correspond to degradation of

the full length protein. No bands were detected in any samples from the empty vector purification suggesting the protein contaminants do not possess His-tag-like motifs.

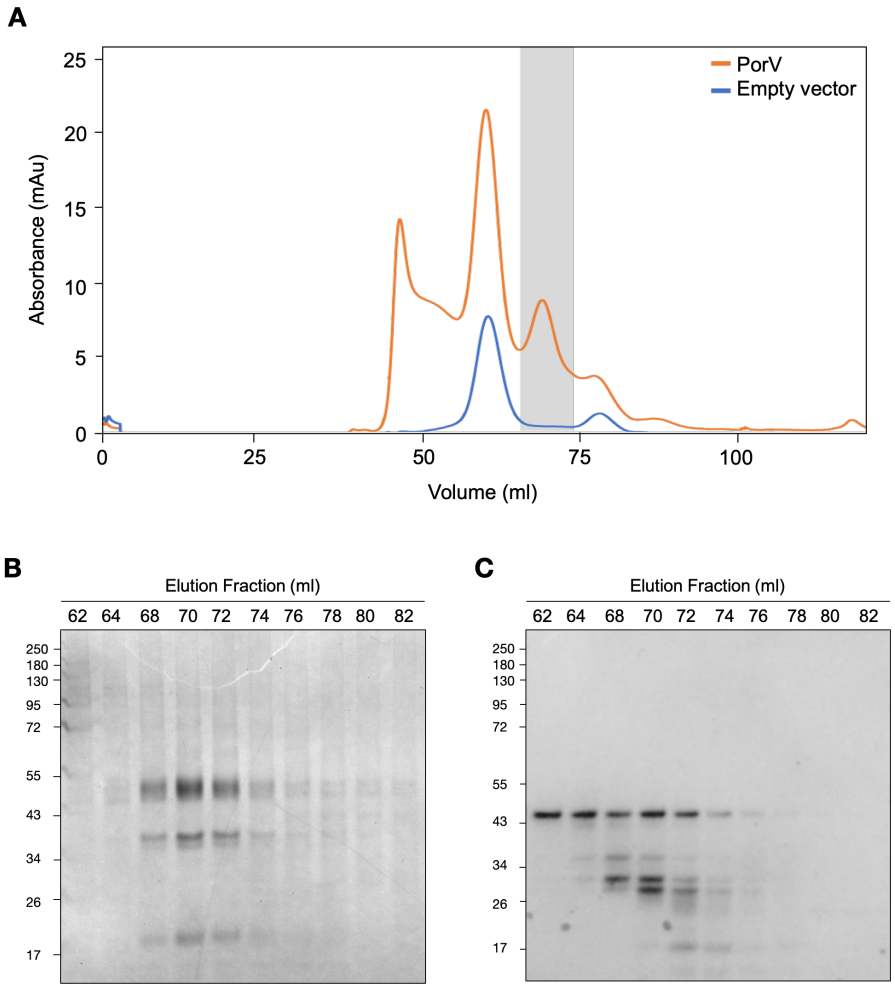


Figure 6.3.6: Result of size exclusion chromatography for His-PorV. (A) UV A280 absorbance chromatogram the His-PorV and pET28(b)-OSS empty vector purification. The His-PorV peak is highlighted in grey. (B) SDS-PAGE and anti-His-HRP western blot (C) of the His-PorV 2 ml fractions from 62 ml to 82 ml.

6.3.7 Oligomerisation analysis

Based on the observed elution volume of His-PorV using size exclusion chromatography the predicted molecular weight of this peak is ~130 kDa. A number of factors could result in this observed disparity including potential oligomerisation of the PorV

protein. Previous analysis of the PorU:Z:V:Q attachment complex by mass spectrometry showed the molar ratios of these proteins was 1:1:3:1, respectively [112]. This is consistent with a trimeric oligomerisation state for PorV. The expected molecular weight for a PorV trimer, in complex with a LDAO micelle, is ~130 kDa.

Therefore, the purified His-PorV protein treated using the reversible chemical cross-linker DSP (120 µM)s with samples collected every 10 minutes for a total of 60 minutes. The samples were separated by SDS-PAGE, under non-reducing conditions, and analysed using anti-His-HRP western blot (6.3.7). This showed that there is no propensity for PorV to form higher oligomerisation states following the addition of DSP, suggesting that PorV is monomeric in this purification.

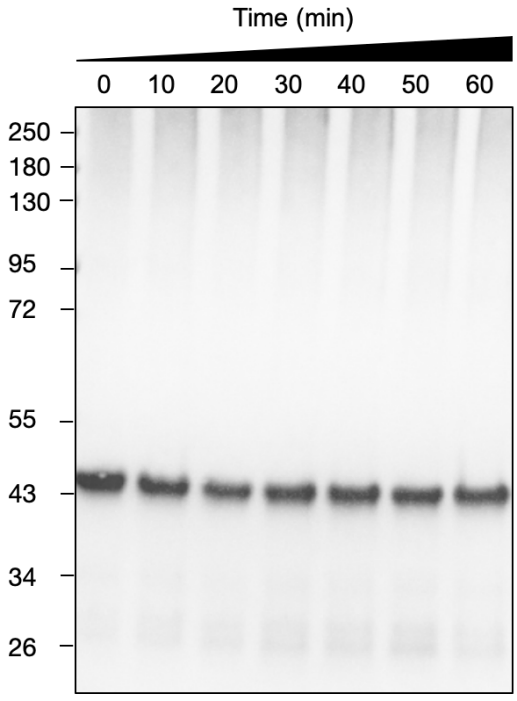


Figure 6.3.7: Oligomerisation analysis of His-PorV using DSP cross-linking. A concentration of 120 µM DSP was used with samples taken at 10 minute intervals. There was no observed oligomerisation of the PorV protein.

6.3.8 Mass Spectrometry Identification

To confirm that the purified protein sample corresponds to the His-PorV construct a fraction of the purified sample was resolved using SDS-PAGE and sent for analysis by high resolution Orbitrap tandem mass spectrometry coupled to liquid chromatography at the Centre of Excellence for Mass Spectrometry (KCL).

Raw mass spectrometry data were processed into peak list files using Proteome Discoverer V2.2 (ThermoFisher Scientific). The data was searched using the Mascot search algorithm V2.6.0 (MatrixScience) and the Sequest search algorithm [243] against the Uniprot All Taxonomy database (561,911 entries) and Uniprot E.coli BL21(DE3) strain database (4,172 entries). The All Taxonomy database search was performed to identify Bovine Trypsin, used for peptide digestion, and other possible contaminants introduced in the sample processing. There was no identification of PorV in this sample. The Uniprot E. *coli* BL21(DE3) database was then searched to determine the protein contaminants arising from these cells.

Given that PorV could not be identified in either of these databases a unique database containing only the PorV protein sequence was used. Assignment of the peptides showed that PorV is present in the sample with 12 unique peptides and 48% sequence coverage. Comparison to the Uniprot E. *coli* BL21(DE3) strain database search shows that PorV has the highest normalised total spectrum score. In addition to PorV a total of 47 unique proteins from E. *coli* BL21(DE3) were identified.

For low resolution biochemical assays this level of purity was assumed to be sufficient. However, for structural analysis of PorV additional purification stages, or alterations to the existing method, are likely to be necessary.

6.4 *In Silico* Analysis of RgpB-CTD:PorV Complex

Due to lab closures work on the improving the PorV purification and PorV:RgpB-CTD pull-down experiments could not be continued. Therefore, an alternative methodology was developed to generate the necessary structural models and restraints for a model complex of the PorV:RgpB-CTD interaction as experimental restraints, from site-directed mutagenesis, could not be defined. This method involved the generation of a predicted structural model for *P. gingivalis* PorV and the use of molecular dynamic (MD) simulations to sample different conformations of the protein. Representative models from the simulation were then taken for use in docking experiments with RgpB-CTD.

6.4.1 Prediction of *P. gingivalis* PorV Structural Model

During the course of the experimental work in this chapter a cryo-EM model of PorV from *F. johnsoniae* (Fjoh_1555), in complex with the Sov homolog SprA (Fjoh_1653), was published [113]. This structural model of PorV shows that it is a β-barrel protein comprised of 14 β-strands with the C-terminus located at the inner face and the N-terminus located at the outer face of the lipid bilayer. Given that the structural model of PorV is derived from *F. johnsoniae* and the structural model of RgpB-CTD is derived from *P. gingivalis* W50 it is possible that these proteins would not interact due to differences between the two organisms. Therefore, a predicted structural model of PorV from *P. gingivalis* W50 was generated for use in protein docking experiments using the ROBETTA server using comparative modelling with the cyro-EM model of PorV as a structural template. To avoid confusion the *P. gingivalis* predicted PorV model shall be referred to as Pg-PorV and the *F. johnsoniae* experimental model shall

be referred to as Fj-PorV. The quality of the predicted structure was assessed using PROCHECK and PROSA. 99.1% of the dihedrals were located within the favoured and allowed regions and the protein had a Z-score of -3.69. Comparison between Fj-PorV and Pg-PorV, which have 46% sequence identity, showed that they are well aligned with a TM score of 0.84796 and an RMSD of 2.38Å(Figure 6.4.1). Based on the protein quality scores and alignment with the experimental PorV model the Pg-PorV model was considered suitable for further analysis by MD simulations and protein-protein docking.

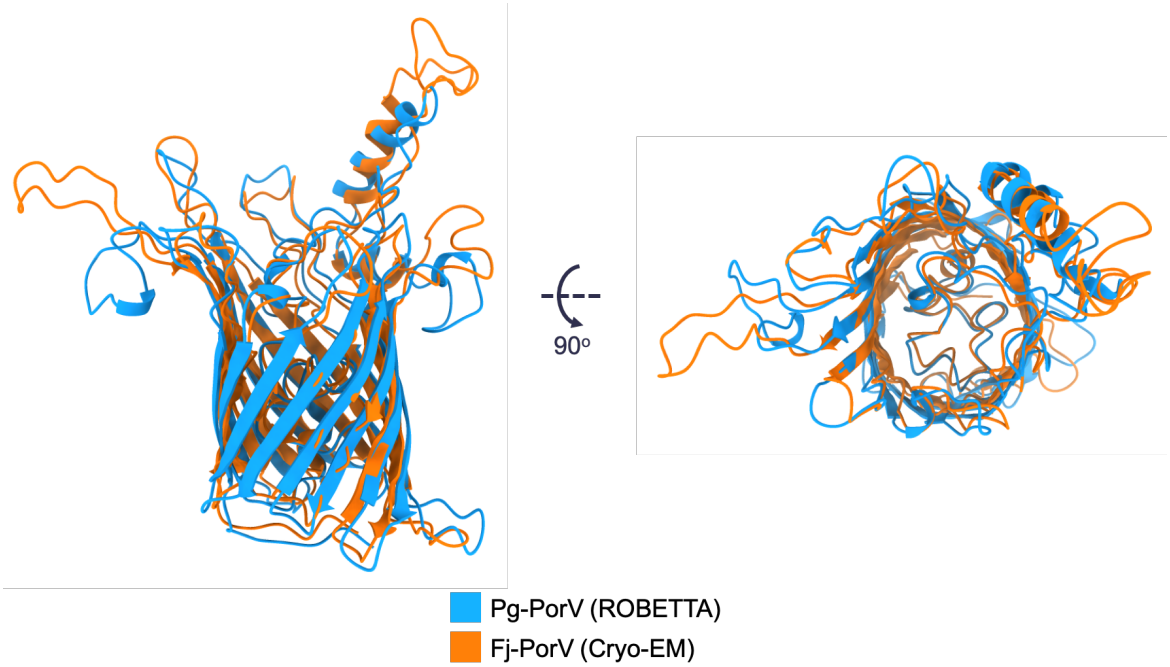


Figure 6.4.1: Comparison of PorV structural models. The cryo-EM model of Fj-PorV (orange) aligned to the ROBETTA model of Pg-PorV (blue). The structures are well aligned with a a TM score of 0.84796 and an RMSD of 2.38Å.

6.4.2 Molecular Dynamics of PorV

Given that the structural model was taken from Fj-PorV in complex with SprA, molecular dynamics (MD) was used to determine how this protein would behave in an unbound state. The second reason for running MD on this protein was to generate

an ensemble of structural models which could be used for computational docking experiments between PorV and RgpB-CTD. Both PorV models were simulated to allow comparison between the experimental and predicted models of PorV.

Orientation of the PorV models was first determined using the online PPM server [195] and uploaded to the CHARMM-GUI membrane builder server [197] for construction of the protein:membrane complex and the necessary input files for MD simulation. The membrane was constructed using 80% POPC and 20% POPE with a total of 100 lipids in each leaflet. Similar lipids compositions are commonly used in the formation of nano-discs and MD simulation of β-barrel proteins from Gram-negative bacteria [244–246]. Given the degree of heterogoneity in *P. gingivalis* LPS composition, and the unavailability of these molecules in the CHARMM-GUI membrane builder, no LPS was included in this simulation. Furthermore, precise values for the relative ratios of each LPS molecule were not available so this component was excluded to avoid artificially biasing protein-LPS interactions in the simulation. GROMACS was used to equilibrate the system to a constant pressure of 1 bar and a temperature of 310 K then energy minimised. Production runs were then performed for a total of 200 ns. Assistance in running and analysing the GROMACS MD simulations was kindly provided by Dr. Chris Lorenz.

6.4.3 Structural Drift and Fluctuation

Analysis of structural drift of the protein from the initial structural model can be used to analyse the stability and motion of the protein during the course of the simulation. This analysis was performed by calculating the RMSD of each frame relative to the first frame (Figure 6.4.2A). In both simulations the RMSD sharply rose for the first 5 ns then reached a steady plateau between 75-100 ns. Both models of PorV showed similar RMSD values during the entire course of the simulation (averages in Table

6.4.1).

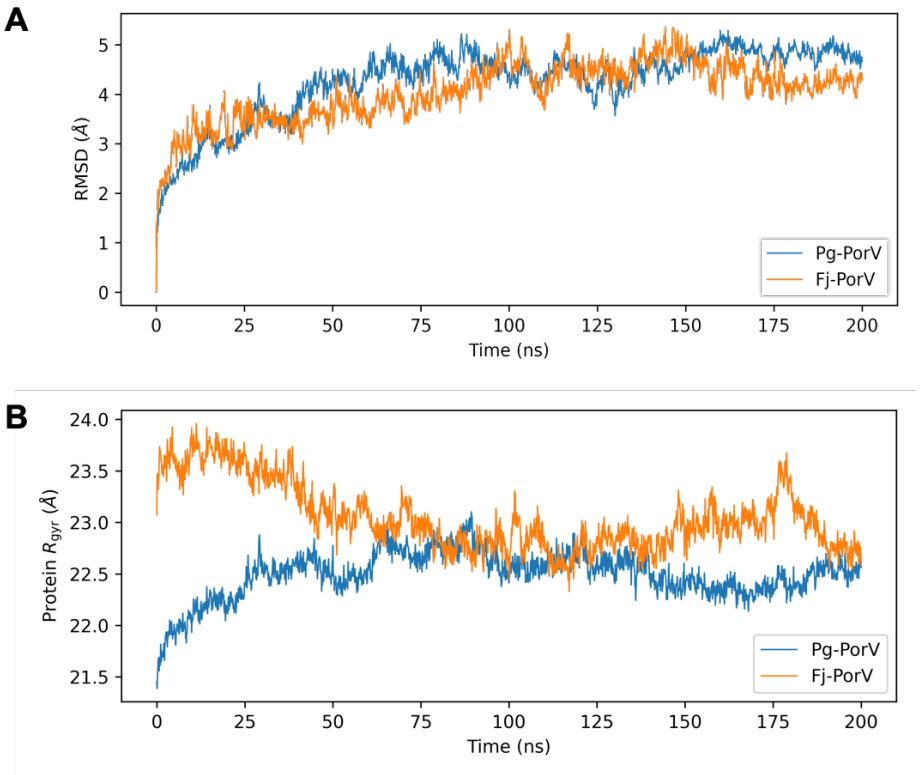


Figure 6.4.2: RMSD of Fj-PorV and Pg-PorV. The protein was fitted to the appropriate starting conformation prior to the MD simulation.

Radius of gyration (RGYR) was calculated for both proteins as a means of approximating protein compactness. The RGYR of fj-PorV decreases within the first 100 ns of the simulation to converge with the RGYR of Pg-PorV, which increases within the first 100 ns of the simulation (Figure 6.4.2B). The two models have comparable RGYR values (averages in Table 6.4.1) with the Fj-PorV model displaying a slightly higher degree of flexibility.

Root mean square fluctuation (RMSF) was used to study how the conformation of protein Cα atoms changes throughout the course of the simulation. As shown in Figure 6.4.3 the β-barrel remains in the same conformation during the course of the simulation but the extracellular loops of PorV show a higher degree of flexibility. The two systems have similar RMSF profiles for the S1-2 and S3-4 loops which are the two

Table 6.4.1: Average RMSD, RGYR and RMSF for Fj-PorV and Pg-PorV.

Measurement	Fj-PorV	Pg-PorV
RMSD (Å)	4.05 ± 0.61	4.25 ± 0.77
RGYR (Å)	23.05 ± 0.32	22.49 ± 0.23
RMSF (Å)	2.00 ± 1.85	1.73 ± 1.67

loops of PorV that enter the SprA pore in the cyro-EM model. Additionally, Pg-PorV showed a higher degree of flexibility in the C-terminus and Fj-PorV displayed a higher degree of flexibility in the S9-S10 loop (residues 263-290). This observed difference could because the Fj-PorV sequence is 18 residues in the S9-S10 loop. The average RMSF for all residues of each model is displayed in Table 6.4.1 and this again suggests that the Fj-PorV model is slightly more flexible than the Pg-PorV model.

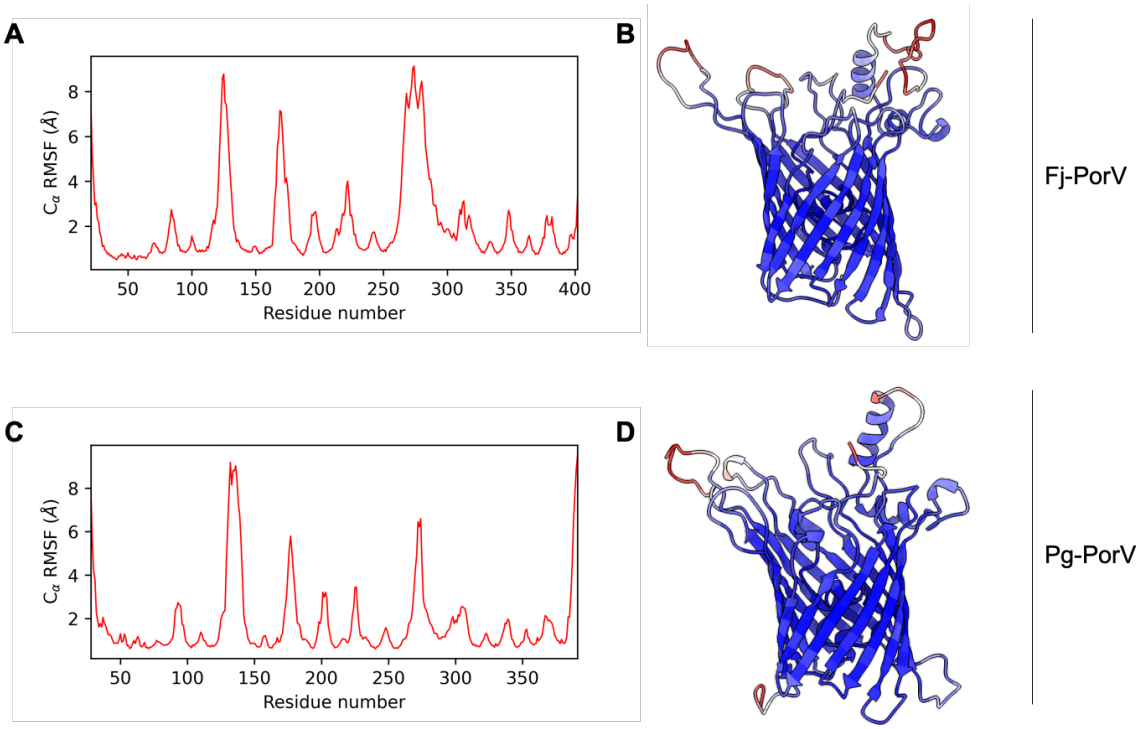


Figure 6.4.3: RMSF of Fj-PorV and Pg-PorV. Plot of RMSF values for Fj-PorV (A) and Pg-PorV (C) Cα atoms in the protein relative to the average structure. The structure of Fj-PorV (B) and Pg-PorV (D) coloured by the RMSF values (blue-red).

6.4.4 Structural Clustering

In order to generate a structural ensemble for use in protein docking experiments the python package TTClust [199] was used to cluster the structures in each frame according to protein Cα RMSD. A linkage matrix is created and the clusters divided according to a user defined cutoff. For Fj-PorV a total of nine clusters was selected using an approximately unbiased (AU) p-value cutoff of 25 (Figure 6.4.4A) and a total of ten clusters were generated for Pg-PorV using an AU cutoff of 22 (Figure 6.4.4B). The RMSD spread of each cluster for Fj-PorV (Figure 6.4.4C) and Pg-PorV (Figure 6.4.4D) was calculated and TTClust used to output a representative model for each cluster. The representative models are those that have the lowest average RMSD all other models within the cluster. The clusters were all of a similar size and with a similar RMSD spread suggesting there no clear preference for a single conformation.

Superposition of the ensemble models confirms that the majority of conformational change in the simulation is localised to the extracellular loops (Figure 6.4.5). It can be observed that there is no conformational changes in the β-barrel portion region of the protein for both models. The same differences in the S9-S10 loop and protein C-terminus flexibility, as observed by RMSF analysis, can be visualised in the ensembles.

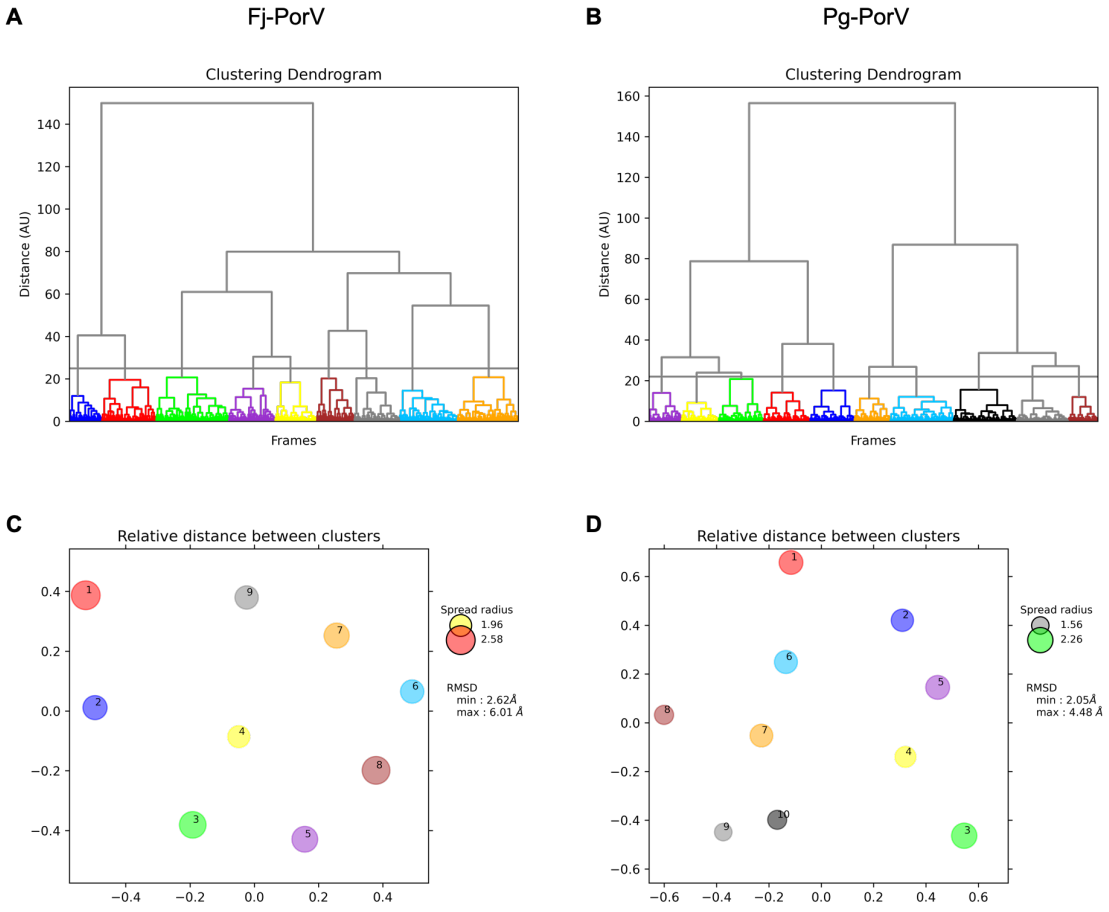


Figure 6.4.4: TTClust analysis of Fj-PorV and Pg-PorV. Dendrogram of Fj-PorV (A) and Pg-PorV (B) clustering. The horizontal grey line represents the cutoff for cluster division. 2D projection of the relative distances between Fj-PorV clusters (C) and Pg-PorV clusters (D) using representative model RMSDs. The minimum and maximum RMSD distances and RMSD spreads are also listed.

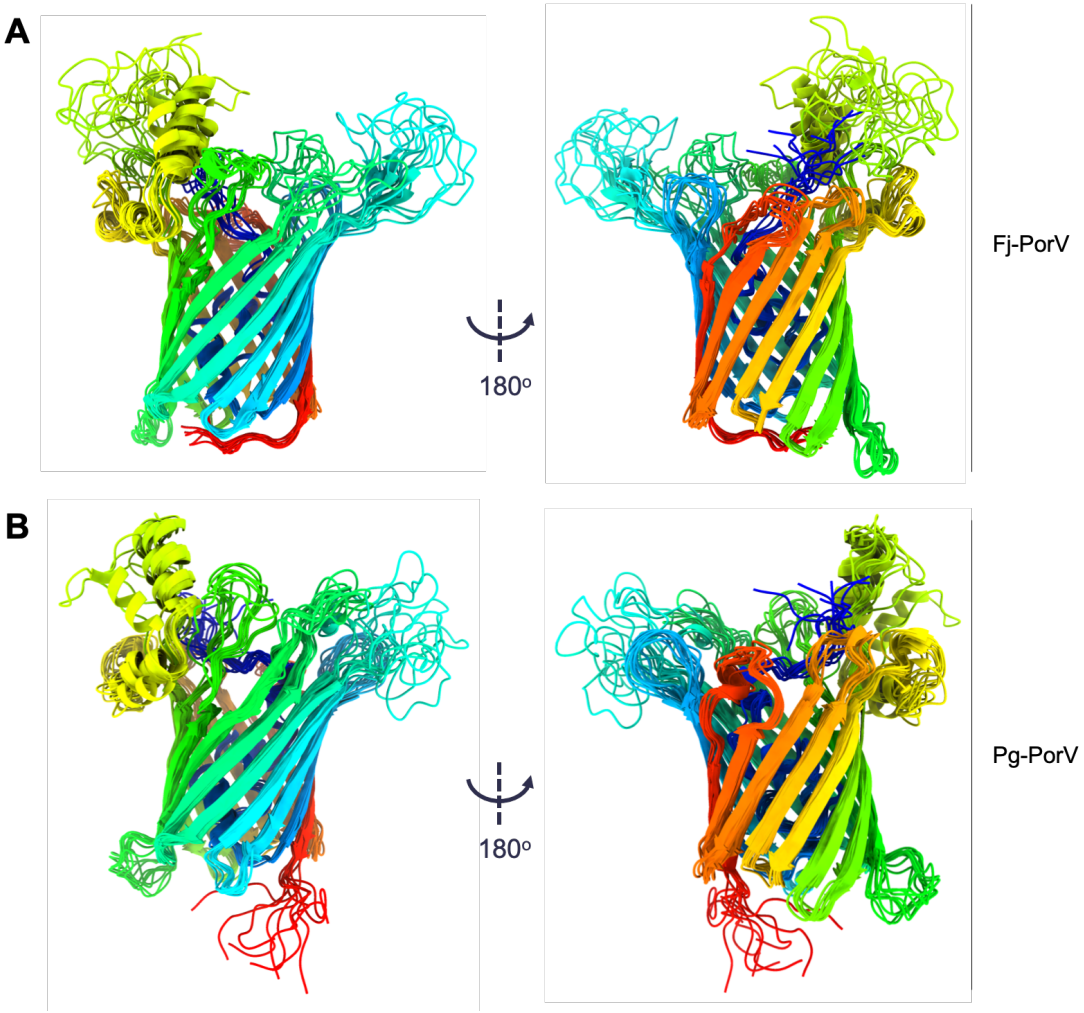


Figure 6.4.5: Superposition of TTClust models. (A) Fj-PorV ensemble and (B) Pg-PorV ensemble show the highest degree of flexibility in the extracellular loops. relative to the Fj-PorV ensemble Pg-PorV has higher flexibility at the protein C-terminus (red).

6.4.5 PorV:RgpB Docking

Comparison of the MD simulations for the *F. johnsoniae* and *P. gingivalis* PorV models showed that they behave similarly within this system. Therefore, it was deemed that the Pg-PorV model was suitable for use in RgpB-CTD:PorV docking experiments. The cluster ensemble of 10 representative models of *P. gingivalis* PorV, defined using TTClust, was selected for the docking experiments. To reduce the computational expense of running docking experiments a single structural model of RgpB-CTD was taken from the ensemble of 10 water-refined models. This model was selected by aligning all models using the online POSA server [201] and the average RMSD for each model to all others was calculated. The structural model with the lowest average RMSD value (1.43 Å) was taken as a representative model of the ensemble.

Molecular docking between the representative RgpB-CTD model and the Pg-PorV ensemble was performed using the online HADDOCK 2.4 server [203]. This algorithm uses ambiguous interaction restraints (AIRs) to define the interaction interface between the two proteins and drive protein docking. The AIRs can be active, essential for the interaction, or passive, involved but not essential. The proteins are then randomly orientated and rigid-body docking with energy minimisation performed. AIRs are important at this stage as they are included in the energy minimisation function. An MD based refinement is then used to introduce flexibility to interface residues (those with < 5Å contact between proteins) to further optimise the interface. A short MD run is then carried out for each complex solvated in water to further refine the structure.

No experimental restraints were available for this interaction so predicted restraints were used. To produce the necessary AIRs, the RgpB-CTD model and PorV ensemble were uploaded to the CPORT server [202]. This server uses a combination of six interface prediction servers to generate a consensus interaction interface prediction

and is specifically designed for use with the HADDOCK algorithm. It generates a list of active and passive AIRs to be used in HADDOCK docking. For RgpB-CTD the CPORT active residue list was used as the HADDOCK active restraint list and the passive residues were defined automatically. For the PorV ensemble the CPORT active residue restraints that appeared in at least half of the models were used selected and the list reduced to residues are localised beyond the outer leaflet of the membrane. This was done because many predicted interacting residues in PorV were membrane embedded. Figure 6.4.6 shows the RgpB-CTD model and an example PorV model coloured by CPORT residue prediction. The predicted RgpB-CTD interface residues are localised to the C-terminal face of the domain and contains the conserved motifs B, D and E. The reduced list of interface residues in Pg-PorV are localised to loops S1-2 and S3-4 which corresponds to the loops of Fj-PorV that enter the SprA pore.

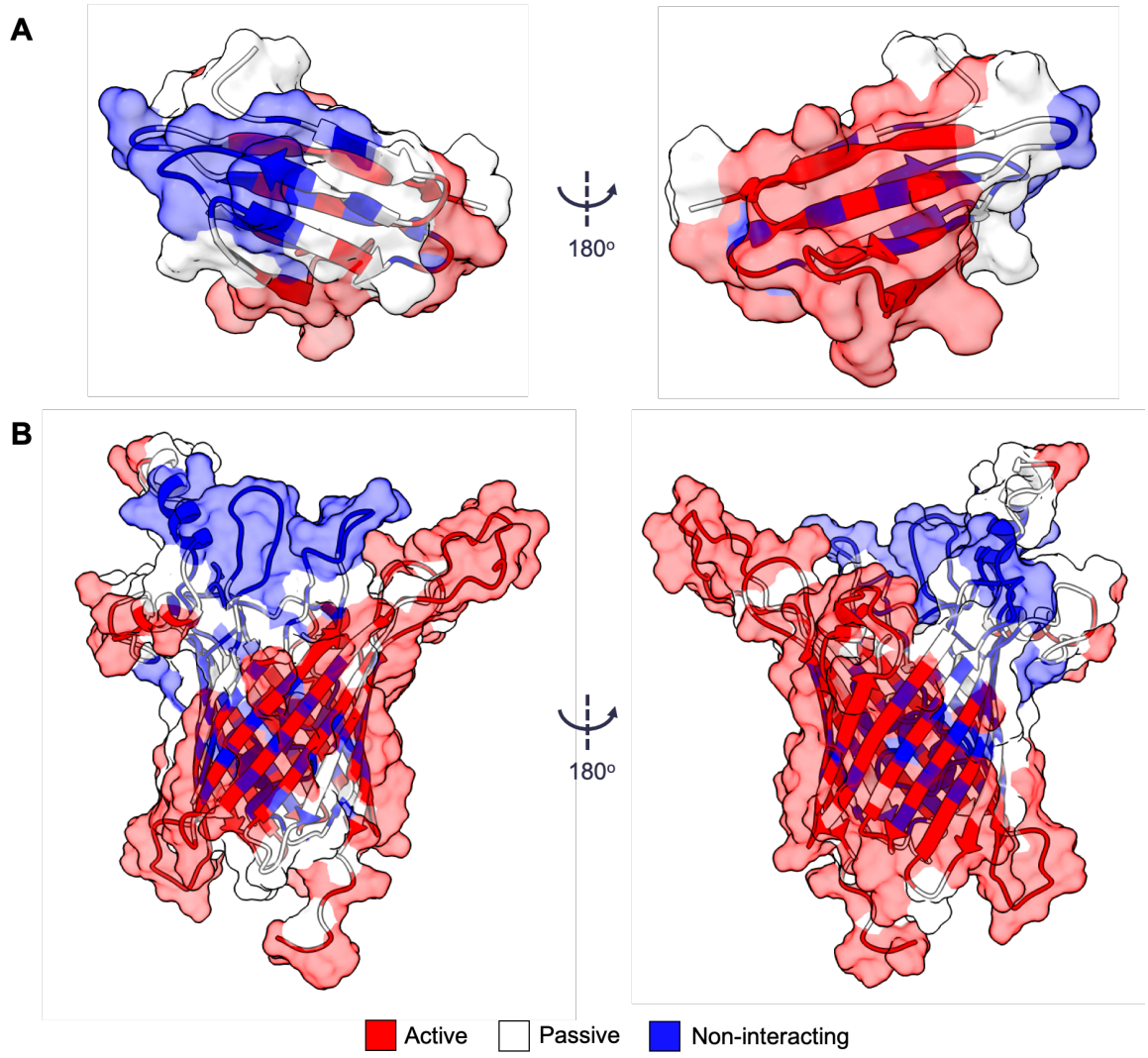


Figure 6.4.6: CPORT analysis of PorV and RgpB-CTD structural models. (A) Representative RgpB-CTD model coloured by CPORT prediction. (B) Example model of PorV coloured by CPORT prediction. The PorV residue list was further refined to those occurring in at least half the ensemble models and localised beyond the outer leaflet of the membrane. Predicted active residues are highlighted in red, passive residues are highlighted in white and non-interacting residues are highlighted in blue.

Of the 200 final water refined structures generated by HADDOCK a total of 9 clusters were produced, comprised of 63 structural models (31%). These models were first sorted according to their HADDOCK score [203]. This score is calculated for each structural complex after docking and is used to rank the quality of the structures produced in a given run. The score is a weighted sum of the intermolecular electrostatics, van der Waals energy, desolvation energy, AIRs energy and buried surface area.

In addition, the interface RMSD (IRMSD) was calculated for the 4 best scoring structures within each cluster. IRMSD is calculated for the backbone atoms of all residues within 10Åof the intermolecular contact compared to the top ranking structure. According to the membrane protein docking benchmark for HADDOCK 2.4 [247], an IRMSD of $2-4$Å is considered acceptable, $1-2$Å is considered medium and <1Å is considered high quality. Therefore, all clusters with an IRMSD >4Å were excluded from further analysis. This left five clusters with an IMRSD above the cutoff limit. Of these, cluster 3 was also excluded because the RgpB-CTD was interfacing with membrane embedded residues of PorV. According to these criteria cluster 5 is high quality, cluster 2 is medium quality and clusters 1, 7 and 6 are acceptable. The run statistics for the lowest scoring four structures of the refined cluster list is shown in Table 6.4.2.

Figure 6.4.7 shows the binding poses for the best scoring model of the five clusters with no clear convergence to a particular location. However, there is a preference for the CTD to bind to the upper/inner face of the S3-4 loop. It is only in cluster 1 that the CTD binds to the lower/outer face of the S3-4 loop.

Table 6.4.2: Statistics for top 4 scoring structures of each HADDOCK cluster.

Parameter	**Cluster 2**	**Cluster 5**	**Cluster 1**	**Cluster 7**	**Cluster 6**
HADDOCK Score	$-116/9 \pm 9.7$	-115.6 ± 19.4	-94.5 ± 5.3	-90.0 ± 15.5	-72.2 ± 15.9
Cluster size	11	5	18	4	4
RMSD* (Å)	4.9 ± 0.9	1.0 ± 0.6	8.3 ± 0.2	4.8 ± 0.1	7.1 ± 0.5
Van der Waals energy (Kcal/mol)	-42.5 ± 4.1	-53.8 ± 8.0	-31.8 ± 9.8	-27.2 ± 13.0	-39.6 ± 5.2
Electrostatic energy (Kcal/mol)	-454.3 ± 28.4	-378.8 ± 56.9	-412.7 ± 59.0	-427.1 ± 64.3	-231.7 ± 20.6
Desolvation energy (Kcal/mol)	7.9 ± 1.1	5.9 ± 4.9	11.4 ± 1.4	11.0 ± 3.9	-0.7 ± 2.3
Restraints violation energy (Kcal/mol)	85.5 ± 14.8	81.1 ± 11.6	85.1 ± 24.7	116.2 ± 60.0	143.9 ± 69.8
Buried surface area (Å)	1788.4 ± 124.4	1756.6 ± 135.5	1422.2 ± 96.9	1509.8 ± 221.5	1520.1 ± 68.2
Z-Score	-1.1	-1.0	0.5	0.8	2.0
IRMSD (Å)	1.165 ± 0.3	0.5824 ± 0.4	2.243 ± 0.3	2.655 ± 0.6	2.403 ± 0.7
TTClust cluster	2	6	7	5	2

* RMSD relative to lowest energy structure in cluster.

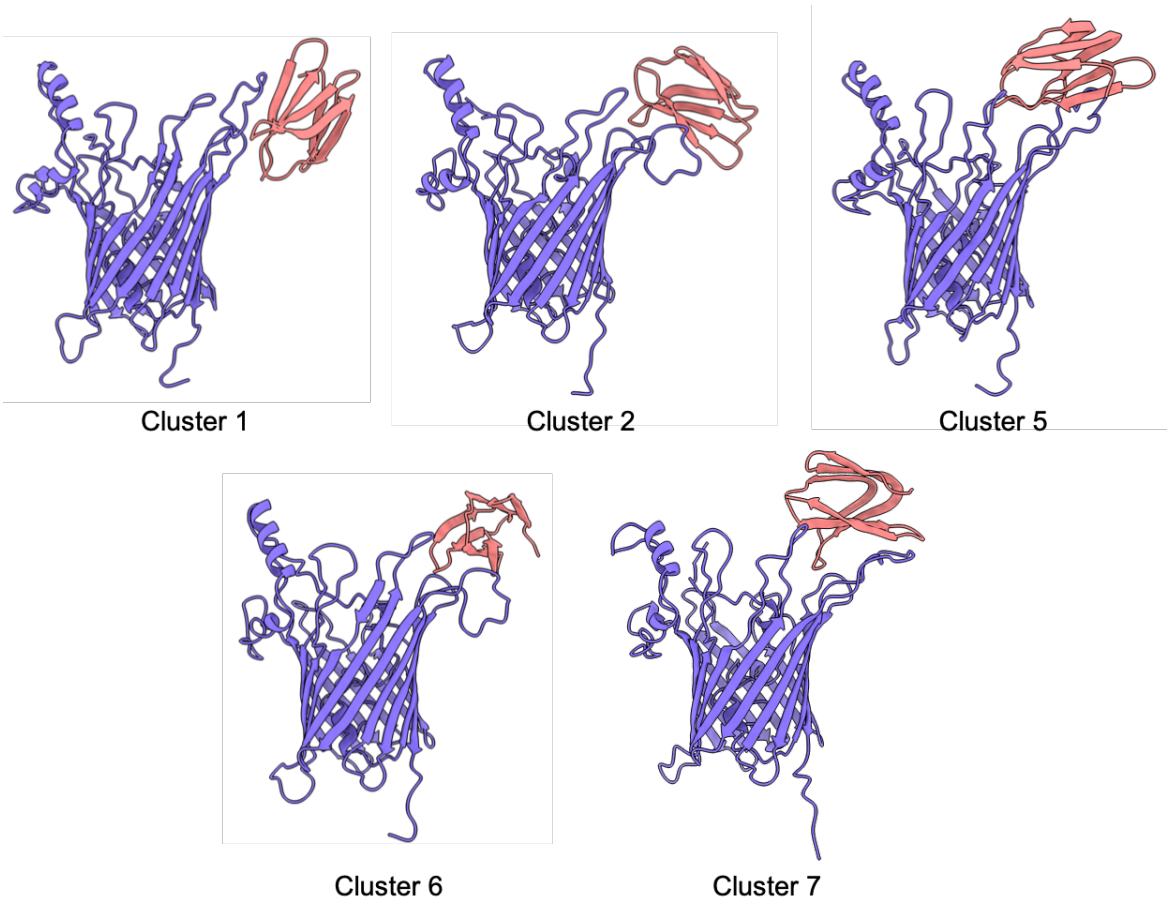


Figure 6.4.7: Binding poses for refined HADDOCK cluster list. PorV is shown in purple an rgpB-CTD is shown in pink. The best scoring structure of each cluster is represented.

Hydrogen bond analysis was performed for the best scoring structure from cluster 5 using the online server COCOMAPS [248]. The results of this analysis are shown in Table 6.4.3.

Consensus inter-residue contacts were calculated using the CONSRANK server [249] for the four best structures of cluster 1, 2, 5, 6 and 7. The rank is calculated from the conversation of each inter-residue contact for all structures. The results of this analysis for contacts with a conservation rate > 0.5 are listed in Table 6.4.4. This analysis revealed that the S5-6 loop of PorV (R^{170}-A^{188}) is participating in the interaction despite no AIRs being selected for residues in this loop. Furthermore, all

Table 6.4.3: H-Bonds between best structure of cluster 5. Atom type is shown in brackets.

PorV Residue	RgpB-CTD Residue	Distance (Å)
E129 (OE1)	K732 (NZ)	2.67
H130 (OG1)	R721 (NH2)	2.79
132D (OD1)	N698 (ND2)	2.81
E133 (OE2)	R701 (NH1)	2.65
L134 (O)	R721 (NH1)	2.82
E136 (O)	R721 (NH2)	2.74
S137(OG)	K732 (NZ)	2.76
M138 (N)	K732 (N)	3.03
E140 (OE2)	K732 (NZ)	3.05
E140 (OE2)	K736 (NZ)	2.96
D172 (OD2)	K736 (NZ)	2.71
S174 (O)	K736 (NZ)	2.86
S174 (OG)	K736 (NZ)	3.92
H176 (NE2)	I734 (O)	2.72
N177 (ND2)	V678 (O)	3.06

residues with a conservation rank above 0.5 are located in loops S3-4 and S5-6 and not in the S1-2 loop. All residues in RgpB-CTD that have a conservation score above 0.5 are located within motifs B, D and E corresponding to the residues predicted using CPORT.

6.4.6 PorV Conservation

An NCBI BlastP [162] search using the *P. gingivalis* W50 PorV sequence, with an expect threshold of 0.05, was used to identify 500 aligned sequence. These were used to generate a multiple sequence alignment for the BlastP results using the Clustal Omega online server [164]. The ConSurf online server [167] was used to colour the

Table 6.4.4: Consensus inter-residue contacts for top four structures. The refined list of clusters 1, 2, 5, 6 and 7 was used.

PorV Residue	RgpB-CTD Residue	Conservation rate
E129	K732	0.65
H176	V735	0.60
G135	T728	0.60
E140	K732	0.60
E133	R721	0.60
L134	T728	0.55
H176	I734	0.55
E136	T728	0.55
M138	T730	0.5
L128	M697	0.5
L128	N698	0.5
E129	A719	0.5

final trajectory model of Pg-PorV according to sequence conservation using the Clustal Omega multiple sequence alignment (Figure 6.4.8A). The multiple sequence alignment was uploaded to the Weblogo server to identify sequence conservation for each residue position. The conservation for the S1-2, S3-4 and S5-6 loops is shown in Figure 6.4.8B, numbered according the W50 RgpB-PorV sequence.

This multiple sequence alignment shows that the sequence in the S1-2 loop is well conserved (Figure 6.4.8B). The five N-terminal residues of the S3-4 loop and P^{148} are well conserved but the central region of the loop is highly variable. S1-2 and the N-terminus of S3-4 form a surface exposed patch of well conserved residues, highlighted in Figure 6.4.8A. The sequence alignment analysis shows that there are some highly conserved residues in the S5-6 loop, particularly D^{172}, whereas middle of the S5-6 sequence is the most variable. As shown in Figure 6.4.8A, it is the variable region of the S5-6 loop that is surface-exposed and accessible for RgpB-CTD binding.

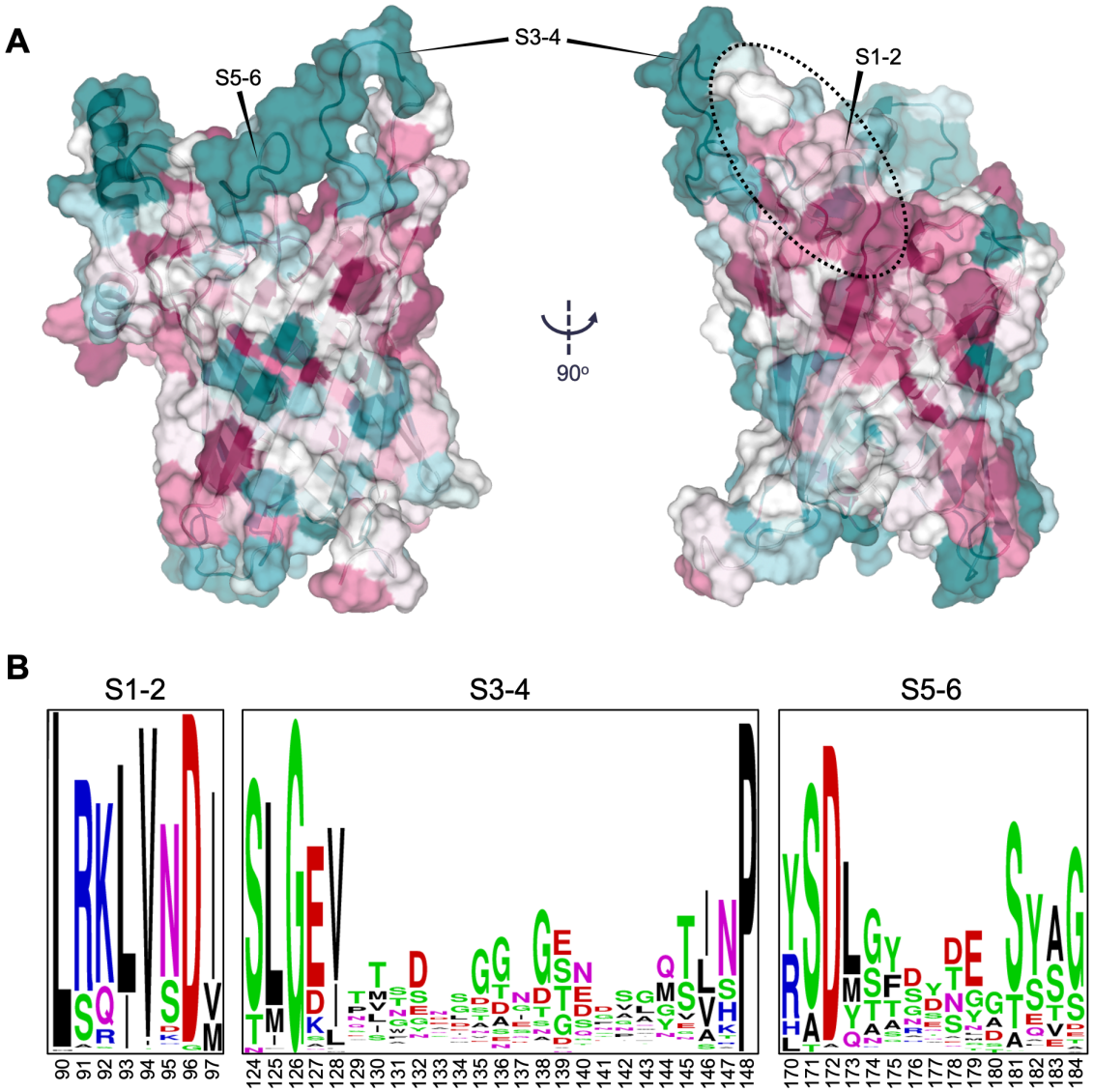


Figure 6.4.8: ConSurf of Pg-PorV using the final the final trajectory structure as a model for the colouring. A patch of conserved residues contained within loops S1-2 and S3-4 is highlighted.

6.5 Discussion

6.5.1 PorV Expression and Purification

In this chapter significant progress has been made towards the expression and purification of *P. gingivalis* PorV. This has not previously been recorded in the literature and access to purified PorV protein sample will provide a useful tool for future experiments exploring the mechanistic process of T9SS cargo protein secretion and processing. The method described here is also a good foundation for further improvement in the expression and purification of recombinant PorV.

The OmpA signal peptide was success at targeting the fusion protein to the outer membrane and the His-Tag allowed for successful purification of PorV protein. However, the protein yield could be further optimised and sample purity requires improvement. Analysis of the protein sample UV absorbance (A_{280}) showed that the protein yield following size exclusion chromatography is ~0.1mg $ml^{-1}$for 1 L of culture. This yield is far below the amount required for NMR spectroscopy but could be sufficient for small scale crystallography plate screening. However, before beginning the process of plate screening the sample purity requires further refinement. This is because the impurities in the sample may lead to crystallography artefacts that give poorly diffracting crystals or the contaminant proteins may form crystals more readily than PorV [250].

As expected the majority of contaminant proteins identified by LC-MS were of a similar size to the PorV construct. However, the third highest scoring protein in the LC-MS was Cytochrome ubiquinol oxidase subunit 1 (~74 kDa). This protein is a known contaminant of LDAO purifications from *E. coli*, along with Cytochrome ubiquinol oxidase subunit 2 (~35 kDa). No record in the literature was found of the two highest scoring *E. coli* proteins (Elongation factor Tu and L-lactate dehydrogenase)

occurring as contaminants in *E. coli* protein purifications. The two peaks seen either side of the PorV peak in Figure 6.3.6 could correspond to contaminants identified by LC-MS or could be different proteins entirely. An important next step in improving the PorV purification method is to identify the proteins comprising the contamination peaks. This knowledge can then be used to determine the best method of contaminant removal.

The detergent screen revealed that LDAO is significantly better at solubilising PorV compared to the other solubilising agents tested. However, the sample purity is poor following Ni-NTA chromatography and changing the detergent using for protein solubilisation may increase sample purity. To do this with any of the detergents tested here would significantly decrease to protein yield. However, a more extensive screen of different detergents could help to address this. Alternatively, transfer to a different detergent after the initial solubilisation may assist in contaminant removal. If sample purity could be improved still using LDAO for solubilisation and purification this detergent may be sufficient for future crystallisation as this detergent has been used to crystallise the homologous membrane protein FadL [251].

More extensive screening of expression systems and tags could be also performed to improve the sample quality. The original Phyre2 model of PorV (Figure 6.2.2) was created prior to the publication of the cryo-EM model of PorV from *F. johnsoniae* [113]. Comparison of the two models shows that the initial prediction of N- and C-termini orientation was correct. It is therefore not clear from the positioning of the protein termini why the C-terminally tagged constructs formed inclusion bodies and the N-terminally tagged construct successfully integrated into the outer membrane. Use of a different N-terminal tag may improve the purification process due to different non-specific binding patterns. It is interesting to note that cytochrome C oxidase is also a known contaminant of Strep-tag II purifications for *E. coli* (MerkBioscience).

DSP cross-linking of PorV suggests that the construct purified here is monomeric. This may be a result of the purification method used here or it could be that PorV is monomeric in the bacterial membrane. The cryo-EM complex of the SprA-PorV would suggest that porV is monomeric in the membrane as only a single PorV molecule is present in this complex [113]. PorV may be forming oligomers as part of the attachment complex only and may require the presence of PorU, Q, Z and/or cargo protein to initiate oligomerisation.

6.5.2 PorV Molecular Dynamics

The Fj-PorV model was simulated alongside the Pg-PorV model to ensure that the predicted model was behaving sensibly within the system and producing valid models that could be used for protein docking. The results of this analysis show that the behaviour of the two models is similar. The most obvious difference occurs in the flexibility of the S9-S10 loop where Fj-PorV had an additional 18 residues. The density of these residues was missing in the cryo-EM structure of Fj-PorV supporting the observation of high flexibility. Whether this structural difference conveys any functional difference between the two organisms is unclear. Interestingly, for some of Pg-PorV models used in the CPORT server run some residues in the S9-S10 loop were predicted to be active protein interface residues. This loop could therefore be facilitating other protein interactions, potentially in the formation of the attachment complex. This could be investigated further in future studies.

In both models the extracellular loops showed increased flexibility relative to the rest of the protein. This, along with the interface prediction from CPORT, would suggests that it is the S1-S2, S3-S4 and S5-6 that interact with Sov to capture CTD proteins in *P. gingivalis* as is the case in *F. johnsoniae* [113]. The increased flexibility of these loops may play a role in facilitating their entry into the Sov channel and allow

for conformational rearrangement as the cargo is pulled through the pore.

However, the flexibility observed here could, in part, be due to the exclusion of any LPS component during the simulation. Cryo-EM of *P. gingivalis* cells and OMVs shows that they a possess an electron-dense surface layer (EDSL) of ~20 nm [108]. This layer is predicted to be composed of LPS molecules and attached T9SS cargo proteins [84]. The presence of this layer would provide additional spacial constraints of the extracellular PorV loops and specific interactions between the EDSL and PorV may occur. This could limit the degree of flexibility observed during the MD simulation using a lipid bilayer only.

6.5.3 PorV:RgpB-CTD Interaction

The interface residues identified by the CPORT server correspond well with the region of PorV that enters the SprA channel in the cryo-EM model. Despite, the S5-S6 loop entering the SprA channel, no interface residues were predicted for this loop in the Pg-PorV model. The majority of interfacing residues predicted in RgpB-CTD by CPORT are located in conserved motifs B, D and E and correspond to conserved residues identified in Section 4.2.1.

In the resulting HADDOCK dock for Pg-PorV and RgpB-CTD, the involvement of the conserved CTD motif B (I^{694}-V^{703}) and motif D (A^{713}-A^{723}) is clear (Table 6.4.4). This supports the theory that the last two strands of the CTD (motif D and E) are essential for cargo protein recognition by the T9SS [88]. Despite the difficulty in identifying motif E in the multiple sequence alignment shown in Section 4.2.1, residues K^{732} and V^{735}were shown to be highly conserved. Both of these residues were found to be consensus interfacing residues in the RgpB-CTD:Pg-PorV interaction (Table 6.4.4). Residue T^{728} was shown to be a consensus interface residue but is not well conserved either in W50 CTDs or across species containing RgpB. Therefore, this residue would a

interesting target for further experimental analysis to determine if it plays a significant role in the interaction.

The previously published $D^{696}A$ mutant, for RgpB in *P. gingivalis*, showed secretion of RgpB but no surface attachment. It was therefore suggested that motif B does not play a role in cargo protein secretion but may provide a signal for cargo processing. In the HADDOCK model presented here the moderately conserved residues M^{697} and N^{698} are both consensus interfacing residues, suggesting motif B does appear to play a role in the binding of PorV to the CTD. Perhaps this motif is not essential for recognition of the cargo in Sov but helps to maintain cargo cell surface attachment via the PorV interaction.

The docking structures suggest that the S5-S6 loop is involved in the interaction despite no predicted residues being identified by CPORT. This interaction is consistent with current understanding of the system as the S5-6 residues would be able to interface with the CTD present in the Sov channel. However, sequence conservation analysis of this loop reveals that the exposed residues are not well conserved. The interfacing between S5-6 and RgpB-CTD could be an artefact of the docking simulation due to the close proximity of this loop to the S1-2 and S3-4 binding site. Alternatively, the S5-6 loop may only play an accessory role in the interaction and may not be essential for the interaction.

Despite different binding poses being observed in the five clusters the orientation of RgpB-CTD relative to PorV was the same in each cluster. As shown by the interaction conservation scoring (Table 6.4.4) the S1-2 loop of PorV interacts primarily with motif B, whereas motifs D and E interface primarily with the S3-4 loop with secondary interactions to the S5-6 loop. The consensus binding residues from PorV are predominantly located within the non-conserved region of the S3-4 loop. This could be due to differences in cargo protein recognition between different species or could be due to

incorrect docking in the HADDOCK simulations. Experimental validation (Discussed in Section 7.2.3) of the residues involved in the interaction could help to determine if the non-conserved residues of the S3-4 loop are essential for the Pg-PorV:RgpB-CTD interaction .

The hydrogen bond analysis was performed to determine the involvement of electrostatic in the interaction. This is a low resolution model given that a predicted structure and interface restraints were used. However, it does suggest that electrostatic interactions play a significant role in the interface with charge complementarity between the two proteins. As shown in Table 6.4.3 the majority of H-bond donors are present in RgpB-CTD and the majority of acceptors are from Pg-PorV.

6.6 Summary

This chapter describes the design of a recombinant PorV construct for use *in vitro* interaction studies. A method of expression and purification was developed and optimised to yield purified and folded PorV protein. The final refinement of the purification method and progression to *in vitro* interaction studies with RgpB-CTD was halted due to lab closures. A predicted model of Pg-PorV was generated using the cryo-EM model of F-PorV and MD simulation was ran for both proteins in a lipid bilayer system. This was used to generate an ensemble of Pg-PorV models sampling different conformations. Predicted interaction residues were predicted for Pg-PorV and the NMR structural model of RgpB-CTD. The partner proteins were docked using the HADDOCK 2.4 algorithm to produce structural models of the complex showing interaction between motifs B, D and E of RgpB-CTD with loops S1-2, S3-4, S5-6 of PorV. The docking models can now be used for the design of point mutations to give experimental restraints to the interaction.

Chapter 7

Conclusions and Future Perspectives

7.1 Conclusions

The work in this project has sought to understand how the CTD of T9SS cargo proteins can facilitate secretion. RgpB-CTD was selected for analysis and successfully expressed and purified. This thesis contains the first NMR solution-state structural model of a T9SS cargo protein CTD showing it to be a seven stranded β-sandwich domain. The RgpB-CTD NMR structural model has been compared to the recently published X-ray crystallography models of RgpB-, Hbp35- and PorZ-CTDs [92, 170, 225]. Analysis of the structures has confirmed that the CTD is structurally conserved and that sequentially conserved motifs B, D and E are located at the C-terminal face of the domain. This has provided insight into the important structural motifs of the CTD that facilitate secretion and suggest that it is the C-terminal face that is recognised by the system. This has been supported by the CPORT interface prediction that suggests all interfacing residues are located at the C-terminal face.

The NMR relaxation data presented here suggest= that the domain is monomeric

in solution. Furthermore, the X-ray crystallography structural models for RgpB-CTD containing the preceding IgSF domain, and the structural models of Hbp35 and PorZ are also monomeric. The size of the SprA pore in the Cryo-EM model also suggests that cargo proteins are secreted as monomers [113]. This evidence supports the conclusion that RgpB-CTD is a monomer in solution state and is recognised by the T9SS as a monomeric domain.

Work was undertaken to identify P. gingivalis proteins that interact with the CTD of RgpB. Unfortunately, the method described here was not successful in identifying components of the T9SS that interact with RgpB-CTD. It is clear now that working with PorV would not have yielded positive results as this would have caused the plug protein to close the Sov channel [113]. However, it is unclear why the PorV:CTD complex could not be captured from Wt *P. gingivalis* using the Co-immunoprecipitation method. Given that the antibody used was polyclonal it is expected that there would be solvent-accessible epitopes in RgpB-CTD if the current HADDOCK docking model is correct.

Despite being unsuccessful in observing any interactions using the non-specific screening method, evidence of an interaction between PorV and T9SS cargo proteins was published [112] but no direct binding to the CTD was demonstrated. Therefore, the PorV protein was selected for direct interaction studies with RgpB-CTD. Significant progress has been made towards the recombinant expression and purification of this protein. The current method is capable of providing purified PorV for biochemical assays that do not require highly purified protein samples, such as protein pull-down assays. For an experimental structural model to be developed the expression and purification method requires alteration to improve protein yield and purity. The data presented here, along with the SprA:PorV structure [113], would suggest that PorV is a monomer in the outer membrane and binds to a single cargo protein. PorV may

then form oligomers as part of the attachment complex with a necessary minimum of two copies: one to bind PorU and one to bind the cargo.

Significant progress has been made in understanding the interaction between RgpB-CTD and PorV. The predicted PorV structural model presented here appeared to align well to the experimental PorV structure from *F. johnsoniae*. The molecular dynamics simulation of this protein suggests that the extracellular loops are highly flexible and capable of sampling different conformations. The docking models generated using representative models of these conformations show that a variety of binding poses are possible.

Figure 7.1.1 shows the two possible mechanisms of secretion according to the binding poses observed in the HADDOCK structures. According to the HADDOCK docking, the CTD can bind to either the lower or upper face of the PorV loops that enter the Sov pore. In the context of protein recognition by the Sov:PorV complex, clusters 2, 5, 6 and 7, in which RgpB-CTD binds to the upper face of the PorV loops, allow for the simplest model of cargo recognition. This is because the N-terminus points into the centre of the Sov channel, towards the periplasm, in these models. In cluster 1, where the CTD is binding to the lower face of PorV, the location of the N-terminus means that the CTD is located 'up-side-down' in this cluster. In all the models the CTD is held above the lipid bilayer and would allow for the cargo to be recognised by and transferred to the attachment complex.

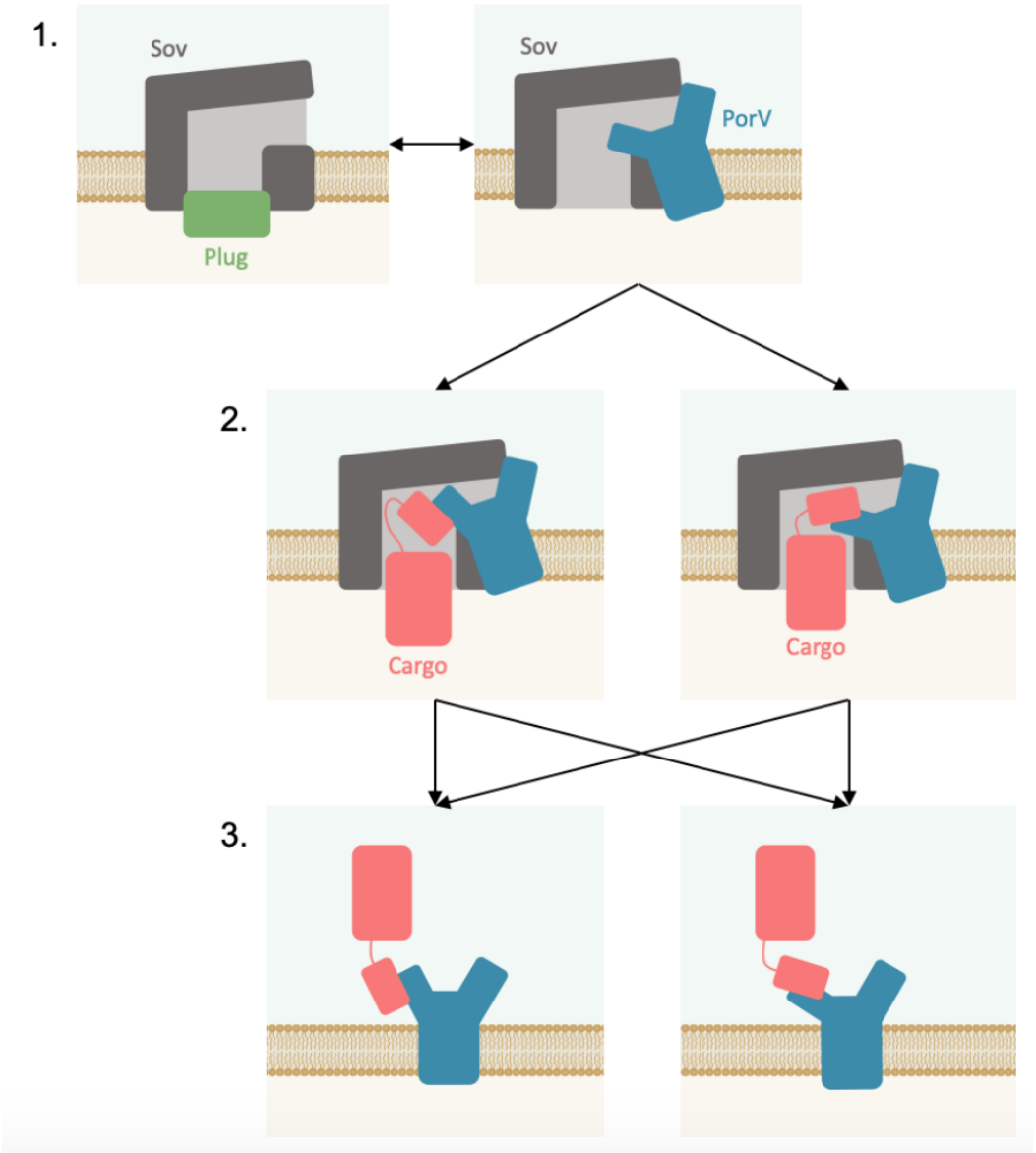


Figure 7.1.1: Possible models of cargo protein recognition by PorV. (1) Sov is either in a plug-bound state, preventing secretion, or in a PorV-bound, allowing for cargo protein recognition. (2) The cargo protein enters the Sov channel and binds to either the upper-face or the lower-face of the PorV loops. (3) The PorV:Cargo complex dissociates from Sov to allow tracking to the attachment complex. It is possible that the cargo stays in the same orientation relative to PorV or a conformational change may occur.

7.2 Future Perspectives

7.2.1 Structural Analysis of Cargo Protein CTDs

The chemical shift assignment for RgpB-CTD can now be used for future interaction studies with T9SS components or for structural analysis of the CTD protein using NMR spectroscopy. This is a valuable tool for probing cargo protein recognition and processing by the T9SS. In the future, it would be beneficial to develop a structural model of the T9SS sortase PorU (Section 1.9) or of the PorU-CTD. This is because the protein is responsible for the removal of cargo protein CTDs and A-LPS attachment, but it is itself a cargo protein of the T9SS. However, the CTD of PorU is not cleaved off and no A-LPS modification occurs. Having a structural model for this protein would allow comparative studies to be performed with the aim of understanding the differences between A-LPS modified cargo proteins and the PorU sortase. By studying the secretion and processing of PorU, and other T9SS cargo proteins, it will be possible to better understand what structural components are conserved or vary between different CTDs.

7.2.2 Identification of CTD Interaction Partners

Despite the identification of many T9SS components that are essential for cargo protein recognition and processing [98], functional characterisation for most of these proteins is lacking. There is a great deal of interest in assigning a functional role of these proteins in the context of the T9SS. One aspect of this, that this thesis has attempted to address, is how cargo proteins are recognised, secreted and processed by the T9SS. Although the non-specific screening approach employed here was unsuccessful, this type of screening would be beneficial in future studies for the identification of CTD interaction partners. Use of alternative T9SS component mutants that do not cause

the plug protein to block the Sov pore could be identified and used in the screening method. For example, a ΔPorM mutant may remove the T9SS motor, preventing export of cargo proteins, but may still allow cargo proteins to bind to other system components. Using this method may yield positive results and provide a pool of potential interaction partners that can be experimentally examined in further detail.

7.2.3 PorV:RgpB-CTD Structural Complex

To generate experimental interaction restraints for the HADDOCK docking, work has already started to directly demonstrate the interaction between recombinant PorV and RgpB-CTD using protein pull-down assays. Once this method has been developed, and if the interaction is verified, it will be possible to apply this to mutant constructs of PorV and RgpB-CTD. The published conservation data [81], previous point mutation studies [207, 226], the CPORT prediction and the CONSRANK interface conservation analysis can be used in future studies for the rational design of mutations in PorV and RgpB-CTD. Given that PorV loops S1-2, S3-4 and S5-6 all appear to be involved in the interaction, mutations should target residues in all three loops. This would help to clarify the role of the S5-6 loop in the interaction given lack of sequence conservation for this loop. The same method can also be used for conserved residues identified in RgpB-CTD. Based on the interaction profile of these mutants, new active restraints can be defined for future HADDOCK runs.

Docking experiments may also be performed using the Hbp35-CTD models, and any future published CTD models, to allow cross-examination in CTD recognition by PorV. This could validate the importance of certain structural motifs in cargo protein recognition. It appears that PorQ may directly bind to Sov to facilitate PorZ transport as PorZ is still found in the outer membrane fraction in a PorU deletion mutant [112]. How this is mediated and regulated is still unknown and further research in this area

is required.

By generating an experimentally verified structural complex of PorV:RgpB-CTD we will better understand how cargo proteins bind to and are processed by the T9SS. In the future, it would be beneficial to develop complexes for CTD proteins with components of the T9SS at different stages of secretion. This will allow us to understand which structural components of the CTD are essential for all stages of secretion and what structural components guide secretion at a specific stage only.

7.2.4 Two Types of CTD

There is an increasing body of evidence that some cargo proteins are secreted, the CTD removed and A-LPS attached to the cargo as a surface anchor. However, not all cargos follow this process. PorU and PorZ both possess a T9SS CTD but are anchored to the cell surface via interaction with PorV and PorQ, respectively. Furthermore, PG1035 was found to bind to PorP at the cell surface and had an intact CTD (preprint: [89]). The CTD of PG1035 was found to be similar to the 'type B' CTD found in *F. johnsoniae*. In *F. johnsoniae* there are 11 type B CTD proteins with 11 corresponding PorP-like proteins [252]. The disparity in processing between different CTD proteins is an interesting area that requires further investigation. Questions also exist regarding the commonality of the CTD, given that all cargo proteins appear to pass through the Sov channel. Further study of the different CTD structures and their interaction with the T9SS should begin to resolve these questions regarding the secretion of cargo proteins.

References

[1] How, K. Y., Song, K. P. & Chan, K. G. (2016). *Porphyromonas gingivalis*: An overview of periodontopathic pathogen below the gum line. *Frontiers in Microbiology*, 7(FEB): 1–14.

[2] Mysak, J., Podzimek, S., Sommerova, P., Lyuya-Mi, Y., Bartova, J., Janatova, T., Prochazkova, J. & Duskova, J. (2014). *Porphyromonas gingivalis*: Major periodontopathic pathogen overview. *Journal of Immunology Research*, 2014: 1–8.

[3] Dietrich, T., Ower, P., Tank, M., West, N. X., Walter, C., Needleman, I., Hughes, F. J., Wadia, R., Milward, M. R., Hodge, P. J. & Chapple, I. L. C. (2019). Periodontal diagnosis in the context of the 2017 classification system of periodontal diseases and conditions – implementation in clinical practice. *British Dental Journal*, 226(1): 16–22.

[4] Kinane, D. F., Stathopoulou, P. G. & Papapanou, P. N. (2017). Periodontal diseases. *Nature Reviews Disease Primers*, 3: 1–14.

[5] Frencken, J. E., Sharma, P., Stenhouse, L., Green, D., Laverty, D. & Dietrich, T. (2017). Global epidemiology of dental caries and severe periodontitis – a comprehensive review. *Journal of Clinical Periodontology*, 44(S18): S94–S105.

[6] Eke, P. I., Dye, B. A., Wei, L., Slade, G. D., Thornton-Evans, G. O., Borgnakke, W. S., Taylor, G. W., Page, R. C., Beck, J. D. & Genco, R. J. (2015). Update on prevalence of eriodontitis inadults in the United States: NHANES 2009 to 2012. *Journal of Periodontology*, 86(5): 611–622.

[7] AlJehani, Y. A. (2014). Risk factors of periodontal disease: Review of the literature. *International Journal of Dentistry*, 2014(2014): 1–9.

[8] Alpagot, T., Duzgunes, N., Wolff, L. F. & Lee, A. (2004). Risk factors for periodontitis in HIV+ patients. *Journal of Periodontal Research*, 39(3): 149–157.

[9] Agholme, M. B., Dahllöf, G. & Modéer, T. (1999). Changes of periodontal status in patients with Down syndrome during a 7-year period. *European Journal of Oral Sciences*, 107(2): 82–88.

[10] Dhadse, P., Gattani, D. & Mishra, R. (2010). The link between periodontal disease and cardiovascular disease: How far we have come in last two decades? *Journal of Indian Society of Periodontology*, 14(3): 148–154.

[11] Preshaw, P. M. & Bissett, S. M. (2019). Periodontitis and diabetes. *British Dental Journal*, 227(7): 577–584.

[12] Rodríguez-Lozano, B., González-Febles, J., Garnier-Rodríguez, J. L., Dadlani, S., Bustabad-Reyes, S., Sanz, M., Sánchez-Alonso, F., Sánchez-Piedra, C., González-Dávila, E. & Díaz-González, F. (2019). Association between severity of periodontitis and clinical activity in rheumatoid arthritis patients: A case–control study. *Arthritis Research & Therapy*, 21(1): 27.

[13] Dominy, S. S., Lynch, C., Ermini, F., Benedyk, M., Marczyk, A., Konradi, A., Nguyen, M., Haditsch, U., Raha, D., Griffin, C., Holsinger, L. J., Arastu-Kapur, S., Kaba, S., Lee, A., Ryder, M. I., Potempa, B., Mydel, P., Hellvard, A., Adamowicz, K., Hasturk, H., Walker, G. D., Reynolds, E. C., Faull, R. L. M., Curtis, M. A., Dragunow, M. & Potempa, J. (2019). *Porphyromonas gingivalis* in Alzheimer's disease brains: Evidence for disease causation and treatment with small-molecule inhibitors. *Science Advances*, 5(1): eaau3333.

[14] Fitzpatrick, R. E., Wijeyewickrema, L. C. & Pike, R. N. (2009). The gingipains: Scissors and glue of the periodontal pathogen, *Porphyromonas gingivalis*. *Future microbiology*, 4(4): 471–487.

[15] Jenkinson, H. F. & Lamont, R. J. (2005). Oral microbial communities in sickness and in health. *Trends in Microbiology*, 13(12): 589–595.

[16] Holt, S. C. & Ebersole, J. L. (2005). *Porphyromonas gingivalis*, *Treponema denticola*, and *Tannerella forsythia*: The "red complex", a prototype polybacterial pathogenic consortium in periodontitis. *Periodontology 2000*, 38: 72–122.

[17] Hajishengallis, G., Darveau, R. P. & Curtis, M. A. (2012). The keystone-pathogen hypothesis. *Nature Reviews Microbiology*, 10(10): 717–725.

[18] Lamont, R. J. & Hajishengallis, G. (2015). Polymicrobial synergy and dysbiosis in inflammatory disease. *Trends in Molecular Medicine*, 21(3): 172–183.

[19] Cutler, C. W., Kalmar, J. R. & Genco, C. A. (1995). Pathogenic strategies of the oral anaerobe, *Porphyromonas gingivalis*. *Trends in Microbiology*, 3(2): 45–51.

[20] Mark Welch, J. L., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E. & Borisy, G. G. (2016). Biogeography of a human oral microbiome at the micron scale. *Proceedings of the National Academy of Sciences*, 113(6): E791–E800.

[21] Takahashi, N., Saito, K., Schachtele, C. F. & Yamada, T. (1997). Acid tolerance and acid-neutralizing activity of *Porphyromonas gingivalis*, *Prevotella*

intermedia and *Fusobacterium nucleatum*. *Oral Microbiology and Immunology*, 12(6): 323–328.

[22] Takahashi, N., Sato, T. & Yamada, T. (2000). Metabolic pathways for cytotoxic end product formation from glutamate- and aspartate-containing peptides by *Porphyromonas gingivalis*. *Journal of Bacteriology*, 182(17): 4704–4710.

[23] Takahashi, N. & Sato, T. (2001). Preferential utilization of dipeptides by *Porphyromonas gingivalis*. *Journal of Dental Research*, 80(5): 1425–1429.

[24] Hajishengallis, G., Liang, S., Payne, M. A., Hashim, A., Jotwani, R., Eskan, M. A., McIntosh, M. L., Alsam, A., Kirkwood, K. L., Lambris, J. D., Darveau, R. P. & Curtis, M. A. (2011). Low-abundance biofilm species orchestrates inflammatory periodontal disease through the commensal microbiota and complement. *Cell Host & Microbe*, 10(5): 497–506.

[25] Moore, W. E., Holdeman, L. V., Smibert, R. M., Hash, D. E., Burmeister, J. A. & Ranney, R. R. (1982). Bacteriology of severe periodontitis in young adult humans. *Infection and Immunity*, 38(3): 1137–1148.

[26] Doungudomdacha, S., Rawlinson, A. & Douglas, C. W. (2000). Enumeration of *Porphyromonas gingivalis*, *Prevotella intermedia* and *Actinobacillus actinomycetemcomitans* in subgingival plaque samples by a quantitative-competitive PCR method. *Journal of Medical Microbiology*, 49(10): 861–874.

[27] Kumar, P. S., Leys, E. J., Bryk, J. M., Martinez, F. J., Moeschberger, M. L. & Griffen, A. L. (2006). Changes in periodontal health status are associated with bacterial community shifts as assessed by quantitative 16S cloning and sequencing. *Journal of Clinical Microbiology*, 44(10): 3665–3673.

[28] Moore, W. E., Moore, L. H., Ranney, R. R., Smibert, R. M., Burmeister, J. A. & Schenkein, H. A. (1991). The microflora of periodontal sites showing active destructive progression. *Journal of Clinical Periodontology*, 18(10): 729–739.

[29] Chaves, E. S., Jeffcoat, M. K., Ryerson, C. C. & Snyder, B. (2000). Persistent bacterial colonization of *Porphyromonas gingivalis*, *Prevotella intermedia*, and *Actinobacillus actinomycetemcomitans* in periodontitis and its association with alveolar bone loss after 6 months of therapy. *Journal of Clinical Periodontology*, 27(12): 897–903.

[30] Van Steenbergen, T. J., Delemarre, F. G., Namavar, F. & De Graaff, J. (1987). Differences in virulence within the species *Bacteroides gingivalis*. *Antonie Van Leeuwenhoek*, 53(4): 233–244.

[31] Laine, M. L., Appelmelk, B. J. & van Winkelhoff, A. J. (1997). Prevalence and distribution of six capsular serotypes of *Porphyromonas gingivalis* in periodontitis patients. *Journal of Dental Research*, 76(12): 1840–1844.

[32] Brunner, J., Scheres, N., El Idrissi, N. B., Deng, D. M., Laine, M. L., van Winkelhoff, A. J. & Crielaard, W. (2010). The capsule of *Porphyromonas gingivalis* reduces the immune response of human gingival fibroblasts. *BMC Microbiology*, 10(1): 5.

[33] Singh, A., Wyant, T., Anaya-Bergman, C., Aduse-Opoku, J., Brunner, J., Laine, M. L., Curtis, M. A. & Lewis, J. P. (2011). The capsule of *Porphyromonas gingivalis* leads to a reduction in the host inflammatory response, evasion of phagocytosis, and increase in virulence. *Infection and Immunity*, 79(11): 4533–4542.

[34] Njoroge, T., Genco, R. J., Sojar, H. T., Hamada, N. & Genco, C. A. (1997). A role for fimbriae in *Porphyromonas gingivalis* invasion of oral epithelial cells. *Infection and Immunity*, 65(5): 1980–1984.

[35] Amano, A., Nakagawa, I., Okahashi, N. & Hamada, N. (2004). Variations of *Porphyromonas gingivalis* fimbriae in relation to microbial pathogenesis. *Journal of Periodontal Research*, 39(2): 136–142.

[36] Hamada, N., Watanabe, K., Arai, M., Hiramine, H. & Umemoto, T. (2002). Cytokine production induced by a 67-kDa fimbrial protein from *Porphyromonas gingivalis*. *Oral Microbiology and Immunology*, 17(3): 197–200.

[37] Hiramine, H., Watanabe, K., Hamada, N. & Umemoto, T. (2003). *Porphyromonas gingivalis* 67-kDa fimbriae induced cytokine production and osteoclast differentiation utilizing TLR2. *FEMS microbiology letters*, 229(1): 49–55.

[38] Olsen, I. & Singhrao, S. K. (2018). Importance of heterogeneity in *Porhyromonas gingivalis* lipopolysaccharide lipid A in tissue specific inflammatory signalling. *Journal of Oral Microbiology*, 10(1): 1487,742.

[39] Ogawa, T. & Yagi, T. (2010). Bioactive mechanism of *Porphyromonas gingivalis* lipid A. *Periodontology 2000*, 54(1): 71–77.

[40] Darveau, R. P., Pham, T. T. T., Lemley, K., Reife, R. A., Bainbridge, B. W., Coats, S. R., Howald, W. N., Way, S. S. & Hajjar, A. M. (2004). *Porphyromonas gingivalis* lipopolysaccharide contains multiple lipid A species that functionally interact with both toll-like receptors 2 and 4. *Infection and Immunity*, 72(9): 5041–5051.

[41] Liu, R., Desta, T., Raptis, M., Darveau, R. P. & Graves, D. T. (2008). *P. gingivalis* and *E. coli* lipopolysaccharides exhibit different systemic but similar local induction of inflammatory markers. *Journal of Periodontology*, 79(7): 1241–1247.

[42] Kato, H., Taguchi, Y., Tominaga, K., Umeda, M. & Tanaka, A. (2014). *Porphyromonas gingivalis* LPS inhibits osteoblastic differentiation and promotes

pro-inflammatory cytokine production in human periodontal ligament stem cells. *Archives of Oral Biology*, 59(2): 167–175.

[43] Travis, J., Pike, R., Imamura, T. & Potempa, J. (1997). *Porphyromonas gingivalis* proteinases as virulence factors in the development of periodontitis. *Journal of Periodontal Research*, 32(1 Pt 2): 120–125.

[44] Haurat, M. F., Aduse-Opoku, J., Rangarajan, M., Dorobantu, L., Gray, M. R., Curtis, M. A. & Feldman, M. F. (2011). Selective sorting of cargo proteins into bacterial membrane vesicles. *Journal of Biological Chemistry*, 286(2): 1269–1276.

[45] Gui, M. J., Dashper, S. G., Slakeski, N., Chen, Y.-Y. & Reynolds, E. C. (2016). Spheres of influence: *Porphyromonas gingivalis* outer membrane vesicles. *Molecular Oral Microbiology*, 31(5): 365–378.

[46] Nakao, R., Takashiba, S., Kosono, S., Yoshida, M., Watanabe, H., Ohnishi, M. & Senpuku, H. (2014). Effect of *Porphyromonas gingivalis* outer membrane vesicles on gingipain-mediated detachment of cultured oral epithelial cells and immune responses. *Microbes and Infection*, 16(1): 6–16.

[47] Furuta, N., Takeuchi, H. & Amano, A. (2009). Entry of *Porphyromonas gingivalis* outer membrane vesicles into epithelial cells causes cellular functional impairment. *Infection and Immunity*, 77(11): 4761–4770.

[48] Grenier, D., Mayrand, D. & McBride, B. C. (1989). Further studies on the degradation of immunoglobulins by black-pigmented *Bacteroides*. *Oral Microbiology and Immunology*, 4(1): 12–18.

[49] Aduse-Opoku, J., Muir, J., Slaney, J. M., Rangarajan, M. & Curtis, M. A. (1995). Characterization, genetic analysis, and expression of a protease antigen (PrpRI) of *Porphyromonas gingivalis* W50. *Infection and Immunity*, 63(12): 4744–4754.

[50] Fletcher, H. M., Schenkein, H. A., Morgan, R. M., Bailey, K. A., Berry, C. R. & Macrina, F. L. (1995). Virulence of a *Porphyromonas gingivalis* W83 mutant defective in the prtH gene. *Infection and immunity*, 63(4): 1521–1528.

[51] Kesavalu, L., Holt, S. C. & Ebersole, J. L. (1996). Trypsin-like protease activity of *Porphyromonas gingivalis* as a potential virulence factor in a murine lesion model. *Microbial pathogenesis*, 20(1): 1–10.

[52] De Diego, I., Veillard, F., Sztukowska, M. N., Guevara, T., Potempa, B., Pomowski, A., Huntington, J. A., Potempa, J. & Gomis-Rüth, F. X. (2014). Structure and mechanism of cysteine peptidase gingipain K (Kgp), a major virulence factor of *porphyromonas gingivalis*. *Journal of Biological Chemistry*, 289(46): 32,291–32,302.

[53] Imamura, T., Potempa, J., Tanase, S. & Travis, J. (1997). Activation of blood coagulation factor X by arginine-specific cysteine proteinases (Gingipain-Rs) from *Porphyromonas gingivalis*. *Journal of Biological Chemistry*, 272(25): 16,062–16,067.

[54] Imamura, T., Tanase, S., Hamamoto, T., Potempa, J. & Travis, J. (2001). Activation of blood coagulation factor IX by gingipains R, arginine-specific cysteine proteinases from *Porphyromonas gingivalis*. *Biochemical Journal*, 353(Pt 2): 325–331.

[55] Lantz, M. S., Allen, R. D., Duck, L. W., Blume, J. L., Switalski, L. M. & Hook, M. (1991). Identification of *Porphyromonas gingivalis* components that mediate its interactions with fibronectin. *Journal of Bacteriology*, 173(14): 4263–4270.

[56] Imamura, T., Potempa, J., Pike, R. N., Moore, J. N., Barton, M. H. & Travis, J. (1995). Effect of free and vesicle-bound cysteine proteinases of *Porphyromonas gingivalis* on plasma clot formation: Implications for bleeding tendency at periodontitis sites. *Infection and Immunity*, 63(12): 4877–4882.

[57] DeCarlo, A. A., Windsor, L. J., Bodden, M. K., Harber, G. J., Birkedal-Hansen, B. & Birkedal-Hansen, H. (1997). Activation and novel processing of matrix metalloproteinases by a thiol-proteinase from the oral anaerobe *Porphyromonas gingivalis*. *Journal of Dental Research*, 76(6): 1260–1270.

[58] Ryan, M. E. & Golub, L. M. (2000). Modulation of matrix metalloproteinase activities in periodontitis as a treatment strategy. *Periodontology 2000*, 24(1): 226–238.

[59] Popadiak, K., Potempa, J., Riesbeck, K. & Blom, A. M. (2007). Biphasic effect of gingipains from *Porphyromonas gingivalis* on the human complement system. *The Journal of Immunology*, 178(11): 7242–7250.

[60] Zenobia, C. & Hajishengallis, G. (2015). *Porphyromonas gingivalis* virulence factors involved in subversion of leukocytes and microbial dysbiosis. *Virulence*, 6(3): 236–243.

[61] Baggiolini, M. & Clark-Lewis, I. (1992). Interleukin-8, a chemotactic and inflammatory cytokine. *FEBS Letters*, 307(1): 97–101.

[62] Strieter, R. M., Koch, A. E., Antony, V. B., Fick, R. B., Standiford, T. J. & Kunkel, S. L. (1994). The immunopathology of chemotactic cytokines: The role of interleukin-8 and monocyte chemoattractant protein-1. *The Journal of Laboratory and Clinical Medicine*, 123(2): 183–197.

[63] Mikolajczyk-Pawlinska, J., Travis, J. & Potempa, J. (1998). Modulation of interleukin-8 activity by gingipains from *Porphyromonas gingivalis*: Implications for pathogenicity of periodontal disease. *FEBS letters*, 440(3): 282–286.

[64] Sugawara, S., Nemoto, E., Tada, H., Miyake, K., Imamura, T. & Takada, H. (2000). Proteolysis of human monocyte CD14 by cysteine proteinases (gingipains) from *Porphyromonas gingivalis* leading to lipopolysaccharide hyporesponsiveness. *Journal of Immunology (Baltimore, Md.: 1950)*, 165(1): 411–418.

[65] Tada, H., Sugawara, S., Nemoto, E., Takahashi, N., Imamura, T., Potempa, J., Travis, J., Shimauchi, H. & Takada, H. (2002). Proteolysis of CD14 on human gingival fibroblasts by arginine-specific cysteine proteinases from *Porphyromonas gingivalis* leading to down-regulation of lipopolysaccharide-induced Interleukin-8 production. *Infection and Immunity*, 70(6): 3304–3307.

[66] Bhogal, P. S., Slakeski, N. & Reynolds, E. C. (1997). A cell-associated protein complex of *Porphyromonas gingivalis* W50 composed of Arg- and Lys-specific cysteine proteinases and adhesins. *Microbiology*, 143 (Pt 7): 2485–2495.

[67] Veith, P. D., Talbo, G. H., Slakeski, N., Dashper, S. G., Moore, C., Paolini, R. A. & Reynolds, E. C. (2002). Major outer membrane proteins and proteolytic processing of RgpA and Kgp of *Porphyromonas gingivalis* W50. *Biochemical Journal*, 363(Pt 1): 105–115.

[68] Nakayama, K., Ratnayake, D. B., Tsukuba, T., Kadowaki, T., Yamamoto, K. & Fujimura, S. (1998). Haemoglobin receptor protein is intragenically encoded by the cysteine proteinase-encoding genes and the haemagglutinin-encoding gene of *Porphyromonas gingivalis*. *Molecular Microbiology*, 27(1): 51–61.

[69] Shi, Y., Ratnayake, D. B., Okamoto, K., Abe, N., Yamamoto, K. & Nakayama, K. (1999). Genetic analyses of proteolysis, hemoglobin binding, and hemagglutination of *Porphyromonas gingivalis*. Construction of mutants with a combination of rgpA, rgpB, kgp, and hagA. *The Journal of Biological Chemistry*, 274(25): 17,955–17,960.

[70] Li, N. & Collyer, C. A. (2011). Gingipains from *Porphyromonas gingivalis* — complex domain structures confer diverse functions. *European Journal of Microbiology and Immunology*, 1(1): 41–58.

[71] Kadowaki, T., Nakayama, K., Yoshimura, F., Okamoto, K., Abe, N. & Yamamoto, K. (1998). Arg-gingipain acts as a major processing enzyme for various cell surface proteins in *Porphyromonas gingivalis*. *Journal of Biological Chemistry*, 273(44): 29,072–29,076.

[72] Lee, J. Y., Miller, D. P., Wu, L., Casella, C. R., Hasegawa, Y. & Lamont, R. J. (2018). Maturation of the Mfa1 fimbriae in the oral pathogen *Porphyromonas gingivalis*. *Frontiers in Cellular and Infection Microbiology*, 8: 137.

[73] Holt, S. C., Kesavalu, L., Walker, S. & Genco, C. A. (1999). Virulence factors of *Porphyromonas gingivalis*. *Periodontology 2000*, 20(1): 168–238.

[74] Slakeski, N., Cleal, S. M., Bhogal, P. S. & Reynolds, E. C. (1999). Characterization of a *Porphyromonas gingivalis* prtK that encodes a lysine-specific cysteine proteinase and three sequence-related adhesins. *Oral Microbiology and Immunology*, 14(2): 92–97.

[75] Mikolajczyk, J., Boatright, K. M., Stennicke, H. R., Nazif, T., Potempa, J., Bogyo, M. & Salvesen, G. S. (2003). Sequential autolytic processing activates the zymogen of Arg-gingipain. *Journal of Biological Chemistry*, 278(12): 10,458–10,464.

[76] Glew, M. D., Veith, P. D., Chen, D., Seers, C. A., Chen, Y. Y. & Reynolds, E. C. (2014). Blue native-PAGE analysis of membrane protein complexes in *Porphyromonas gingivalis*. *Journal of Proteomics*, 110: 72–92.

[77] Slakeski, N., Bhogal, P. S., O'Brien-Simpson, N. M. & Reynolds, E. C. (1998). Characterization of a second cell-associated Arg-specific cysteine proteinase of P*Porphyromonas gingivalis* and identification of an adhesin-binding motif involved in association of the prtR and prtK proteinases and adhesins into large complexes. *Microbiology (Reading, England)*, 144 (Pt 6): 1583–1592.

[78] Curtis, M. A., Thickett, A., Slaney, J. M., Rangarajan, M., Aduse-Opoku, J., Shepherd, P., Paramonov, N. & Hounsell, E. F. (1999). Variable carbohydrate modifications to the catalytic chains of the RgpA and RgpB proteases of *Porphyromonas gingivalis* W50. *Infection and Immunity*, 67(8): 3816–3823.

[79] Eichinger, A., Beisel, H.-g. G., Jacob, U., Huber, R., Medrano, F.-j. J., Banbula, A., Potempa, J., Travis, J. & Bode, W. (1999). Crystal structure of gingipain R: An Arg-specific bacterial cysteine proteinase with a caspase-like fold. *The EMBO journal*, 18(20): 5453–5462.

[80] Seers, C. A., Slakeski, N., Veith, P. D., Nikolof, T., Chen, Y. Y., Dashper, S. G. & Reynolds, E. C. (2006). The RgpB C-terminal domain has a role in attachment of RgpB to the outer membrane and belongs to a novel C-terminal-domain family found in *Porphyromonas gingivalis*. *Journal of Bacteriology*, 188(17): 6376–6386.

[81] Veith, P. D., Nor Muhammad, N. A., Dashper, S. G., Likić, V. A., Gorasia, D. G., Chen, D., Byrne, S. J., Catmull, D. V. & Reynolds, E. C. (2013). Protein substrates of a novel secretion system are numerous in the *Bacteroidetes phylum* and have in common a cleavable C-terminal secretion signal, extensive post-translational modification, and cell-surface attachment. *Journal of Proteome Research*, 12(10): 4449–4461.

[82] Glew, M. D., Veith, P. D., Peng, B., Chen, Y. Y., Gorasia, D. G., Yang, Q., Slakeski, N., Chen, D., Moore, C., Crawford, S. & Reynolds, E. C. (2012). PG0026 is the C-terminal signal peptidase of a novel secretion system of *Porphyromonas gingivalis*. *Journal of Biological Chemistry*, 287(29): 24,605–24,617.

[83] Hasegawa, Y., Iijima, Y., Persson, K., Nagano, K., Yoshida, Y., Lamont, R., Kikuchi, T., Mitani, A. & Yoshimura, F. (2016). Role of Mfa5 in Expression of Mfa1 Fimbriae in *Porphyromonas gingivalis*. *Journal of Dental Research*, 95(11): 1291–1297.

[84] Veith, P. D., Chen, Y. Y., Gorasia, D. G., Chen, D., Glew, M. D., O'Brien-Simpson, N. M., Cecil, J. D., Holden, J. A. & Reynolds, E. C. (2014). *Porphyromonas gingivalis* outer membrane vesicles exclusively contain outer membrane and periplasmic proteins and carry a cargo enriched with virulence factors. *Journal of Proteome Research*, 13(5): 2420–2432.

[85] Chen, Y.-Y., Cross, K. J., Paolini, R. A., Fielding, J. E., Slakeski, N. & Reynolds, E. C. (2002). CPG70 is a novel basic metallocarboxypeptidase with C-terminal polycystic kidney disease domains from *Porphyromonas gingivalis*. *The Journal of Biological Chemistry*, 277(26): 23,433–23,440.

[86] Aruni, A. W., Lee, J., Osbourne, D., Dou, Y., Roy, F., Muthiah, A., Boskovic, D. S. & Fletcher, H. M. (2012). VimA-dependent modulation of acetyl coenzyme A levels and lipid A biosynthesis can alter virulence in *Porphyromonas gingivalis*. *Infection and Immunity*, 80(2): 550–564.

[87] Nonaka, M., Shoji, M., Kadowaki, T., Sato, K., Yukitake, H., Naito, M. & Nakayama, K. (2014). Analysis of a Lys-specific serine endopeptidase secreted via the type IX secretion system in *Porphyromonas gingivalis*. *FEMS Microbiology Letters*, 354(1): 60–68.

[88] Shoji, M., Sato, K., Yukitake, H., Kondo, Y., Narita, Y., Kadowaki, T., Naito, M. & Nakayama, K. (2011). Por secretion system-dependent secretion and glycosylation of *Porphyromonas gingivalis* hemin-binding protein 35. *PLoS ONE*, 6(6): 1–12.

[89] Gorasia, D. G., Chreifi, G., Seers, C. A., Butler, C. A., Heath, J. E., Glew, M. D., McBride, M. J., Subramanian, P., Kjær, A., Jensen, G. J., Veith, P. D. & Reynolds, E. C. (2020). In situ structure and organisation of the type IX secretion system. *bioRxiv*, p. 2020.05.13.094771.

[90] Nelson, D., Potempa, J., Kordula, T. & Travis, J. (1999). Purification and characterization of a novel cysteine proteinase (periodontain) from *Porphyromonas gingivalis*. Evidence for a role in the inactivation of human alpha1-proteinase inhibitor. *The Journal of Biological Chemistry*, 274(18): 12,245–12,251.

[91] Gorasia, D. G., Veith, P. D., Chen, D., Seers, C. A., Mitchell, H. A., Chen, Y. Y., Glew, M. D., Dashper, S. G. & Reynolds, E. C. (2015). *Porphyromonas gingivalis* type IX secretion substrates are cleaved and modified by a sortase-like mechanism. *PLoS Pathogens*, 11(9): 1–31.

[92] Lasica, A. M., Goulas, T., Mizgalska, D., Zhou, X., de Diego, I., Ksiazek, M., Madej, M., Guo, Y., Guevara, T., Nowak, M., Potempa, B., Goel, A., Sztukowska, M., Prabhakar, A. T., Bzowska, M., Widziolek, M., Thøgersen, I. B., Enghild, J. J., Simonian, M., Kulczyk, A. W., Nguyen, K.-A., Potempa, J. & Gomis-Rüth, F. X. (2016). Structural and functional probing of PorZ, an essential bacterial surface component of the type-IX secretion system of human oral-microbiomic *Porphyromonas gingivalis*. *Scientific Reports*, 6(August): 37,708.

[93] Sato, K., Yukitake, H., Narita, Y., Shoji, M., Naito, M. & Nakayama, K. (2013). Identification of *Porphyromonas gingivalis* proteins secreted by the Por secretion system. *FEMS Microbiology Letters*, 338(1): 68–76.

[94] Casiano-Colón, A. & Marquis, R. E. (1988). Role of the arginine deiminase system in protecting oral bacteria and an enzymatic basis for acid tolerance. *Applied and Environmental Microbiology*, 54(6): 1318–1324.

[95] Vermilyea, D. M., Ottenberg, G. K. & Davey, M. E. (2019). Citrullination mediated by PPAD constrains biofilm formation in *P. gingivalis* strain 381. *npj Biofilms and Microbiomes*, 5(1): 1–11.

[96] Koziel, J., Mydel, P. & Potempa, J. (2014). The link between periodontal disease and Rheumatoid Arthritis: An updated review. *Current Rheumatology Reports*, 16(3): 408.

[97] Shoji, M., Shibata, Y., Shiroza, T., Yukitake, H., Peng, B., Chen, Y. Y., Sato, K., Naito, M., Abiko, Y., Reynolds, E. C. & Nakayama, K. (2010). Characterization of hemin-binding protein 35 (HBP35) in *Porphyromonas gingivalis*: Its cellular distribution, thioredoxin activity and role in heme utilization. *BMC microbiology*, 10(2010): 152.

[98] Lasica, A. M., Ksiazek, M., Madej, M. & Potempa, J. (2017). The type IX secretion system (T9SS): Highlights and recent insights into its structure and function. *Frontiers in Cellular and Infection Microbiology*, 7(May): 1–19.

[99] Okamoto, K., Nakayama, K., Kadowaki, T., Abe, N., Ratnayake, D. B. & Yamamoto, K. (1998). Involvement of a lysine-specific cysteine proteinase in hemoglobin adsorption and heme accumulation by *Porphyromonas gingivalis*. *The Journal of Biological Chemistry*, 273(33): 21,225–21,231.

[100] Hoover, C. I. & Yoshimura, F. (1994). Transposon-induced pigment-deficient mutants of *Porphyromonas gingivalis*. *FEMS microbiology letters*, 124(1): 43–48.

[101] Simpson, W., Wang, C. Y., Mikolajczyk-Pawlinska, J., Potempa, J., Travis, J., Bond, V. C. & Genco, C. A. (1999). Transposition of the endogenous inser-

tion sequence element IS1126 modulates gingipain expression in *Porphyromonas gingivalis*. *Infection and Immunity*, 67(10): 5012–5020.

[102] Chen, T., Dong, H., Yong, R. & Duncan, M. J. (2000). Pleiotropic pigmentation mutants of *Porphyromonas gingivalis*. *Microbial Pathogenesis*, 28(4): 235–247.

[103] Abaibou, H., Chen, Z., Olango, G. J., Liu, Y., Edwards, J. & Fletcher, H. M. (2001). vimA gene downstream of recA is involved in virulence modulation in *Porphyromonas gingivalis* W83. *Infection and Immunity*, 69(1): 325–335.

[104] Shoji, M., Ratnayake, D. B., Shi, Y., Kadowaki, T., Yamamoto, K., Yoshimura, F., Akamine, A., Curtis, M. A. & Nakayama, K. (2002). Construction and characterization of a nonpigmented mutant of *Porphyromonas gingivalis*: Cell surface polysaccharide as an anchorage for gingipains. *Microbiology (Reading, England)*, 148(Pt 4): 1183–1191.

[105] Sato, K., Sakai, E., Veith, P. D., Shoji, M., Kikuchi, Y., Yukitake, H., Ohara, N., Naito, M., Okamoto, K., Reynolds, E. C. & Nakayama, K. (2005). Identification of a new membrane-associated protein that influences transport/maturation of gingipains and adhesins of *Porphyromonas gingivalis*. *Journal of Biological Chemistry*, 280(10): 8668–8677.

[106] Saiki, K. & Konishi, K. (2007). Identification of a *Porphyromonas gingivalis* novel protein Sov required for the secretion of gingipains. *Microbiology and Immunology*, 51(5): 483–491.

[107] Ishiguro, I., Saiki, K. & Konishi, K. (2009). PG27 is a novel membrane protein essential for a *Porphyromonas gingivalis* protease secretion system. *FEMS Microbiology Letters*, 292(2): 261–267.

[108] Chen, Y. Y., Peng, B., Yang, Q., Glew, M. D., Veith, P. D., Cross, K. J., Goldie, K. N., Chen, D., O'Brien-Simpson, N., Dashper, S. G. & Reynolds, E. C. (2011). The outer membrane protein LptO is essential for the O-deacylation of LPS and the co-ordinated secretion and attachment of A-LPS and CTD proteins in *Porphyromonas gingivalis*. *Molecular Microbiology*, 79(5): 1380–1401.

[109] Sato, K., Naito, M., Yukitake, H., Hirakawa, H., Shoji, M., McBride, M. J., Rhodes, R. G. & Nakayama, K. (2010). A protein secretion system linked to *bacteroidete* gliding motility and pathogenesis. *Proceedings of the National Academy of Sciences*, 107(1): 276–281.

[110] Sato, K. (2011). Por secretion system of *Porphyromonas gingivalis*. *Journal of Oral Biosciences*, 53(3): 187–196.

[111] Heath, J. E., Seers, C. A., Veith, P. D., Butler, C. A., Nor Muhammad, N. A., Chen, Y. Y., Slakeski, N., Peng, B., Zhang, L., Dashper, S. G., Cross, K. J.,

Cleal, S. M., Moore, C. & Reynolds, E. C. (2016). PG1058 is a novel multidomain protein component of the bacterial type IX secretion system. *PLoS ONE*, 11(10): 1–29.

[112] Glew, M. D., Veith, P. D., Chen, D., Gorasia, D. G., Peng, B. & Reynolds, E. C. (2017). PorV is an outer membrane shuttle protein for the type IX secretion system. *Scientific Reports*, 7(1): 8790.

[113] Lauber, F., Deme, J. C., Lea, S. M. & Berks, B. C. (2018). Type 9 secretion system structures reveal a new protein transport mechanism. *Nature*, 564(2018): 77–82.

[114] Kadowaki, T., Yukitake, H., Naito, M., Sato, K., Kikuchi, Y., Kondo, Y., Shoji, M. & Nakayama, K. (2016). A two-component system regulates gene expression of the type IX secretion component proteins via an ECF sigma factor. *Scientific Reports*, 6(1): 23,288.

[115] Gorasia, D. G., Veith, P. D., Hanssen, E. G., Glew, M. D., Sato, K., Yukitake, H., Nakayama, K. & Reynolds, E. C. (2016). Structural insights into the PorK and PorN components of the *Porphyromonas gingivalis* type IX secretion system. *PLoS Pathogens*, 12(8): 1–25.

[116] Taguchi, Y., Sato, K., Yukitake, H., Inoue, T., Nakayama, M., Naito, M., Kondo, Y., Kano, K., Hoshino, T., Nakayama, K., Takashiba, S. & Ohara, N. (2016). Involvement of an Skp-like protein, PGN_0300, in the type IX secretion system of *Porphyromonas gingivalis*. *Infection and Immunity*, 84(1): 230–240.

[117] Vincent, M. S., Canestrari, M. J. M. J., Leone, P., Stathopulos, J., Ize, B. B., Zoued, A., Cambillau, C., Kellenberger, C., Roussel, A. & Cascales, E. (2017). Characterization of the *Porphyromonas gingivalis* type IX secretion trans-envelope PorKLMNP core complex. *Journal of Biological Chemistry*, 292(8): 3252–3261.

[118] Leone, P., Roche, J., Vincent, M. S., Tran, Q. H., Desmyter, A., Cascales, E., Kellenberger, C., Cambillau, C. & Roussel, A. (2018). Type IX secretion system PorM and gliding machinery GldM form arches spanning the periplasmic space. *Nature Communications*, 9(1): 429.

[119] Saiki, K. & Konishi, K. (2010). Identification of a novel *Porphyromonas gingivalis* outer membrane protein, PG534, required for the production of active gingipains. *FEMS Microbiology Letters*, 310(2): 168–174.

[120] Vincent, M. S., Durand, E. & Cascales, E. (2016). The PorX response regulator of the *Porphyromonas gingivalis* PorXY two-component system does not directly regulate the type IX secretion genes but binds the PorL subunit. *Frontiers in Cellular and Infection Microbiology*, 6(August): 96.

[121] Trinh, N. T. T., Tran, H. Q., Van Dong, Q., Cambillau, C., Roussel, A. & Leone, P. (2020). Crystal structure of type IX secretion system PorE C-terminal domain from *Porphyromonas gingivalis* in complex with a peptidoglycan fragment. *Scientific Reports*, 10(1): 7384.

[122] James, R. H., Deme, J. C., Kjaer, A., Alcock, F., Silale, A., Lauber, F., Berks, B. C. & Lea, S. M. (2020). Structure of a proton-powered molecular motor that drives protein transport and gliding motility. *bioRxiv*, p. 2020.05.11.089193.

[123] Shrivastava, A., Johnston, J. J., van Baaren, J. M. & McBride, M. J. (2013). *Flavobacterium Johnsoniae* GldK, GldL, GldM, and SprA are required for secretion of the cell surface gliding motility adhesins SprB and RemA. *Journal of Bacteriology*, 195(14): 3201–3212.

[124] Paramonov, N., Rangarajan, M., Hashim, A., Gallagher, A., Aduse-Opoku, J., Slaney, J. M., Hounsell, E. & Curtis, M. A. (2005). Structural analysis of a novel anionic polysaccharide from *Porphyromonas gingivalis* strain W50 related to Arg-gingipain glycans. *Molecular Microbiology*, 58(3): 847–863.

[125] Saiki, K. & Konishi, K. (2014). *Porphyromonas gingivalis* C-terminal signal peptidase PG0026 and HagA interact with outer membrane protein PG27/LptO. *Molecular Oral Microbiology*, 29(1): 32–44.

[126] Vincent, M. S., Chabalier, M. & Cascales, E. (2018). A conserved motif of *Porphyromonas* Type IX secretion effectors C-terminal secretion signal specifies interactions with the PorKLMN core complex. *bioRxiv*, p. 483123.

[127] Nakayama, K. (2015). *Porphyromonas gingivalis* and related bacteria: From colonial pigmentation to the type IX secretion system and gliding motility. *Journal of Periodontal Research*, 50(1): 1–8.

[128] Bloch, F., Hansen, W. W. & Packard, M. (1946). The Nuclear Induction Experiment. *Physical Review*, 70(7-8): 474–485.

[129] Purcell, E. M., Torrey, H. C. & Pound, R. V. (1946). Resonance absorption by nuclear magnetic moments in a solid. *Physical Review*, 69(1-2): 37–38.

[130] Keeler, J. (2010). *Understanding NMR Spectroscopy*. Wiley & Sons, Chichester, UK, second edition.

[131] Hanson, L. G. (2008). Is quantum mechanics necessary for understanding magnetic resonance? *Concepts in Magnetic Resonance Part A*, 32A(5): 329–340.

[132] Marion, D. (2013). An Introduction to Biological NMR Spectroscopy. *Molecular & Cellular Proteomics*, 12(11): 3006–3025.

[133] Wishart, D. S. (2011). Interpreting protein chemical shift data. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 58(1-2): 62–87.

[134] Bader, R. F., Streitwieser, A., Neuhaus, A., Laidig, K. E. & Speers, P. (1996). Electron delocalization and the Fermi hole. *Journal of the American Chemical Society*, 118(21): 4959–4965.

[135] Fukui, H. (2006). The theory of nuclear spin-spin couplings. *Modern Magnetic Resonance*, (5): 79–83.

[136] Li, F., Lee, J. H., Grishaev, A., Ying, J. & Bax, A. (2015). High accuracy of Karplus equations for relating three-Bond J couplings to protein backbone torsion angles. *ChemPhysChem*, 16(3): 572–578.

[137] Plantenga, T. M., Van Zijl, P. C. M. & Maclean, C. (1982). Studies of quadrupolar and dipolar electric field effects in the NMR spectra of binary mixtures of liquids. *Chemical Physics*, 66(1-2): 1–9.

[138] Overhauser, A. W. (1953). Polarization of nuclei in metals. *Phys. Rev.*, 92(2): 411–415.

[139] Hahn, E. L. (1950). Spin echoes. *Physical Review*, 80(4): 580–594.

[140] Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1979). Investigation of exchange processes by two-dimensional NMR spectroscopy. *The Journal of Chemical Physics*, 71(11): 4546–4553.

[141] Aue, W. P., Bartholdi, E. & Ernst, R. R. (1976). Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *The Journal of Chemical Physics*, 64(5): 2229–2246.

[142] Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)*, 89(3): 496–514.

[143] Grzesiek, S. & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance (1969)*, 96(2): 432–440.

[144] Grzesiek, S. & Bax, A. (1992). An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *Journal of Magnetic Resonance (1969)*, 99(1): 201–207.

[145] Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *Journal of Magnetic Resonance, Series B*, 101(2): 201–205.

[146] Bax, A., Clore, G., Driscoll, P. C., Gronenborn, A. M., Ikura, M. & Kay, L. E. (1990). Practical aspects of proton-carbon-carbon-proton three-dimensional correlation spectroscopy of 13C-labeled proteins. *Journal of Magnetic Resonance (1969)*, 87(3): 620–627.

[147] Peti, W., Griesinger, C. & Bermel, W. (2000). Adiabatic TOCSY for C,C and H,H J-transfer. *Journal of Biomolecular NMR*, 18(3): 199–205.

[148] Loquet, A., Bardiaux, B., Gardiennet, C., Blanchet, C., Baldus, M., Nilges, M., Malliavin, T. & Böckmann, A. (2008). 3D structure determination of the Crh protein from highly ambiguous solid-state NMR restraints. *Journal of the American Chemical Society*, 9(5): 3579–3589.

[149] Cheung, M. S., Maguire, M. L., Stevens, T. J. & Broadhurst, R. W. (2010). DANGLE: A bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of Magnetic Resonance*, 202(2): 223–233.

[150] Shen, Y., Vernon, R., Baker, D. & Bax, A. (2009). *De Novo* protein structure generation from incomplete chemical shift assignments. *Journal of Biomolecular NMR*, 43(2): 63–78.

[151] Güntert, P. (2004). Automated NMR structure calculation With CYANA. In Downing, A. K. (editor), *Protein NMR Techniques*, Methods in Molecular Biology™, pp. 353–378. Humana Press, Totowa, NJ.

[152] Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E. & Nilges, M. (2007). ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, 23(3): 381–382.

[153] Linge, J., O'Donoghue, S. & Nilges, M. (2001). Automated assignment of ambiguous nuclear overhauser effects with ARIA. In *Methods in Enzymology*, volume 339, pp. 71–90. Academic Press.

[154] Rosato, A., Tejero, R. & Montelione, G. T. (2013). Quality assessment of protein NMR structures. *Current Opinion in Structural Biology*, 23(5): 715–724.

[155] Grzesiek, S. (2003). Notes on relaxation and dynamics. In *EMBO Practical Course on NMR*, p. 41. Heidelberg.

[156] Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31(13): 3784–3788.

[157] Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In Walker, J. M. (editor), *The Proteomics Protocols Handbook*, Springer Protocols Handbooks, pp. 571–607. Humana Press, Totowa, NJ.

[158] Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1): 105–132.

[159] Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G. & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4): 420–423.

[160] Bagos, P. G., Liakopoulos, T. D., Spyropoulos, I. C. & Hamodrakas, S. J. (2004). PRED-TMBB: A web server for predicting the topology of β-barrel outer membrane proteins. *Nucleic Acids Research*, 32(Web Server Issue): 400–404.

[161] Bagos, P. G., Liakopoulos, T. D., Spyropoulos, I. C. & Hamodrakas, S. J. (2004). A hidden Markov model method, capable of predicting and discriminating β-barrel outer membrane proteins. *BMC Bioinformatics*, 5: 29.

[162] Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, 36(suppl_2): W5–W9.

[163] Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5): 1792–1797.

[164] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1): 539.

[165] Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, 14(6): 1188–1190.

[166] Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, 19(1): 163–164.

[167] Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. & Ben-Tal, N. (2016). ConSurf 2016: An improved methodology to estimate and visualise evolutionary conservation in macromolecules. *Nucleic Acids Research*, 44(W1): W344–W350.

[168] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10(6): 845–858.

[169] Morrison, K. A., Akram, A., Mathews, A., Khan, Z. A., Patel, J. H., Zhou, C., Hardy, D. J., Moore-Kelly, C., Patel, R., Odiba, V., Knowles, T. J., Javed, M.-u.-H., Chmel, N. P., Dafforn, T. R. & Rothnie, A. J. (2016). Membrane protein extraction and purification using styrene–maleic acid (SMA) copolymer: Effect of variations in polymer structure. *Biochemical Journal*, 473(23): 4349–4360.

[170] De Diego, I., Ksiazek, M., Mizgalska, D., Koneru, L., Golik, P., Szmigielski, B., Nowak, M., Nowakowska, Z., Potempa, B., Houston, J. A., Enghild, J. J., Thogersen, I. B., Gao, J., Kwan, A. H., Trewhella, J., Dubin, G., Gomis-Ruth, F. X., Nguyen, K. A. & Potempa, J. (2016). The outer-membrane export signal of *Porphyromonas gingivalis* type IX secretion system (T9SS) is a conserved C-terminal beta-sandwich domain. *Scientific Reports*, 6(23123): 1–17.

[171] Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. (1991). Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *Journal of Magnetic Resonance (1969)*, 93(1): 151–170.

[172] Kay, L., Keifer, P. & Saarinen, T. (1992). Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *Journal of the American Chemical Society*, 114(26): 10,663–10,665.

[173] Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J., Sørensen, O. W. & Griesinger, C. (1994). A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *Journal of biomolecular NMR*, 4(2): 301–306.

[174] Grzesiek, S., Anglister, J. & Bax, A. (1993). Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization. *Journal of Magnetic Resonance, Series B*, 101(1): 114–119.

[175] Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *Journal of Magnetic Resonance, Series B*, 103(3): 203–216.

[176] Clubb, R. T., Thanabal, V. & Wagner, G. (1992). A constant-time 3-dimensional triple-resonance pulse scheme to correlate intraresidue H-1(N), N-15, and C-13' chemical shifts in N-15-C-13 labeled proteins. *Journal of Magnetic Resonance*, 97(1): 213–217.

[177] Montelione, G. T., Lyons, B. A., Emerson, S. D. & Tashiro, M. (1992). An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *Journal of the American Chemical Society*, 114(27): 10,974–10,975.

[178] Uhrin, D., Uhrinova, S., Leadbeater, C., Nairn, J., Price, N. C. & Barlow, P. N. (2000). 3D HCCH3-TOCSY for resonance assignment of methyl-containing side chains in 13C-labeled proteins. *Journal of Magnetic Resonance*, 142(2): 288–293.

[179] Yamazaki, T., Forman-Kay, J. D. & Kay, L. E. (1993). Two-dimensional NMR experiments for correlating carbon-13β and proton δ/ϵ. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. *Journal of the American Chemical Society*, 115(23): 11,054–11,055.

[180] Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989). Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1β. *Biochemistry*, 28(15): 6150–6156.

[181] Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989). Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *Journal of the American Chemical Society*, 111(4): 1515–1517.

[182] Zuiderweg, E. R. P. & Fesik, S. W. (1989). Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry*, 28(6): 2387–2391.

[183] Cordier, F., Nisius, L., Dingley, A. J. & Grzesiek, S. (2008). Direct detection of N-H$\cdots$O=C hydrogen bonds in biomolecules by NMR spectroscopy. *Nature Protocols*, 3(2): 235–241.

[184] Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. (1994). Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, 33(19): 5984–6003.

[185] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, 6(3): 277–293.

[186] Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics*, 59(4): 687–696.

[187] Maciejewski, M. W., Schuyler, A. D., Gryk, M. R., Moraru, I. I., Romero, P. R., Ulrich, E. L., Eghbalnia, H. R., Livny, M., Delaglio, F. & Hoch, J. C. (2017). NMRbox: A resource for biomolecular NMR computation. *Biophysical Journal*, 112(8): 1529–1534.

[188] Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *Journal of biomolecular NMR*, 8(4): 477–486.

[189] Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). Structure validation by Calpha geometry: Phi, psi and C beta deviation. *Proteins*, 50(3): 437–450.

[190] Schrödinger, LLC (2015). The PyMOL molecular graphics system, version 2.

[191] Fiser, A., Do, R. K. & Sali, A. (2000). Modeling of loops in protein structures. *Protein Science : A Publication of the Protein Society*, 9(9): 1753–1773.

[192] Kim, D. N., Moriarty, N. W., Kirmizialtin, S., Afonine, P. V., Poon, B., Sobolev, O. V., Adams, P. D. & Sanbonmatsu, K. (2019). Cryo_fit: Democratization of flexible fitting for cryo-EM. *Journal of Structural Biology*, 208(1): 1–6.

[193] Ovchinnikov, S., Kamisetty, H. & Baker, D. (2014). Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *eLife*, 3: e02,030.

[194] Kim, D. E., Chivian, D. & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*, 32(Web Server issue): W526–W531.

[195] Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. (2012). OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Research*, 40(Database issue): D370–376.

[196] Jo, S., Kim, T., Iyer, V. G. & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11): 1859–1865.

[197] Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., MacKerell, A. D., Klauda, J. B. & Im, W. (2016). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of Chemical Theory and Computation*, 12(1): 405–413.

[198] Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seyler, S. L., Domański, J., Dotson, D. L., Buchoux, S., Kenney, I. M. & Beckstein, O. (2016). MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations. *Proceedings of the 15th Python in Science Conference*, pp. 98–105.

[199] Tubiana, T., Carvaillo, J.-C., Boulard, Y. & Bressanelli, S. (2018). TTClust: A versatile molecular simulation trajectory clustering program with graphical summaries. *Journal of Chemical Information and Modeling*, 58(11): 2178–2182.

[200] Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1): 33–38.

[201] Li, Z., Natarajan, P., Ye, Y., Hrabe, T. & Godzik, A. (2014). POSA: A user-driven, interactive multiple protein structure alignment server. *Nucleic Acids Research*, 42(Web Server Issue): W240–W245.

[202] de Vries, S. J. & Bonvin, A. M. J. J. (2011). CPORT: A consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PLOS ONE*, 6(3): e17,695.

[203] van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Melquiond, A. S. J., van Dijk, M., de Vries, S. J. & Bonvin, A. M. J. J. (2016). The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology*, 428(4): 720–725.

[204] Seabold, S. & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, pp. 92–96. Austin, Texas.

[205] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F. & van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3): 261–272.

[206] Zhou, X.-Y., Gao, J.-L., Hunter, N., Potempa, J. & Nguyen, K.-A. (2013). Sequence-independent processing site of the C-terminal domain (CTD) influences maturation of the RgpB protease from *Porphyromonas gingivalis*. *Molecular Microbiology*, 89(5): 903–917.

[207] Slakeski, N., Seers, C. A., Ng, K., Moore, C., Cleal, S. M., Veith, P. D., Lo, A. W. & Reynolds, E. C. (2011). C-terminal domain residues important for secretion and attachment of RgpB in *Porphyromonas gingivalis*. *Journal of Bacteriology*, 193(1): 132–142.

[208] Bailey, T. L. & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, 2: 28–36.

[209] Frith, M. C., Saunders, N. F. W., Kobe, B. & Bailey, T. L. (2008). Discovering sequence motifs with arbitrary insertions and deletions. *PLOS Computational Biology*, 4(5): e1000,071.

[210] De Diego, I., Veillard, F. T., Guevara, T., Potempa, B., Sztukowska, M., Potempa, J. & Gomis-Rüth, F. X. (2013). *Porphyromonas gingivalis* virulence factor gingipain RgpB shows a unique zymogenic mechanism for cysteine peptidases. *Journal of Biological Chemistry*, 288(20): 14,287–14,296.

[211] Igboin, C. O., Griffen, A. L. & Leys, E. J. (2009). *Porphyromonas gingivalis* strain diversity. *Journal of Clinical Microbiology*, 47(10): 3073–3081.

[212] Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, 189(1): 113–130.

[213] Bodenhausen, G. & Ruben, D. J. (1980). Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters*, 69(1): 185–189.

[214] Sattler, M., Schleucher, J. & Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 34(2): 93–158.

[215] Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallographica. Section D, Biological Crystallography*, 54(Pt 5): 905–921.

[216] Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry*, 28(23): 8972–8979.

[217] Anthis, N. Protein correlation time calculators. http://nickanthis.com/tools/tau.

[218] Neuhaus, D. & van Mierlo, C. P. M. (1992). Measurement of heteronuclear NOE enhancements in biological macromolecules. A convenient pulse sequence for use with aqueous solutions. *Journal of Magnetic Resonance (1969)*, 100(1): 221–228.

[219] Teichmann, S. A. & Chothia, C. (2000). Immunoglobulin superfamily proteins in Caenorhabditis elegans. *Journal of Molecular Biology*, 296(5): 1367–1383.

[220] Williams, A. F. & Barclay, A. N. (1988). The immunoglobulin superfamily–domains for cell surface recognition. *Annual Review of Immunology*, 6: 381–405.

[221] Zhang, Y. & Skolnick, J. (2005). TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 33(7): 2302–2309.

[222] Liu, Y. & Eisenberg, D. (2002). 3D domain swapping: As domains continue to swap. *Protein Science : A Publication of the Protein Society*, 11(6): 1285–1299.

[223] Tudyka, T. & Skerra, A. (1997). Glutathione S-transferase can be used as a C-terminal, enzymatically active dimerization module for a recombinant protease inhibitor, and functionally secreted into the periplasm of *Escherichia coli*. *Protein Science : A Publication of the Protein Society*, 6(10): 2180–2187.

[224] Niedziela-Majka, A., Rymarczyk, G., Kochman, M. & O\.zyhar, A. (1998). GST-Induced dimerization of DNA-binding domains alters characteristics of their interaction with DNA. *Protein Expression and Purification*, 14(2): 208–220.

[225] Sato, K., Kakuda, S., Yukitake, H., Kondo, Y., Shoji, M., Takebe, K., Narita, Y., Naito, M., Nakane, D., Abiko, Y., Hiratsuka, K., Suzuki, M. & Nakayama, K. (2018). Immunoglobulin-like domains of the cargo proteins are essential for protein stability during secretion by the type IX secretion system. *Molecular Microbiology*, 110(1): 64–81.

[226] Nguyen, K.-a. A., Travis, J. & Potempa, J. (2007). Does the importance of the C-terminal residues in the maturation of RgpB from *Porphyromonas gingivalis* reveal a novel mechanism for protein export in a subgroup of gram-negative bacteria. *Journal of bacteriology*, 189(3): 833–843.

[227] Hoffman, E. A., Frey, B. L., Smith, L. M. & Auble, D. T. (2015). Formaldehyde crosslinking: A tool for the study of chromatin complexes. *Journal of Biological Chemistry*, 290(44): 26,404–26,411.

[228] Toews, J., Rogalski, J. C., Clark, T. J. & Kast, J. (2008). Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions. *Analytica Chimica Acta*, 618(2): 168–183.

[229] Thomassin, J.-L., Moreno, J. S., Guilvout, I., Nhieu, G. T. V. & Francetic, O. (2017). The trans-envelope architecture and function of the type 2 secretion system: New insights raising new questions. *Molecular Microbiology*, 105(2): 211–226.

[230] Tagawa, J., Inoue, T., Naito, M., Sato, K., Kuwahara, T., Nakayama, M., Nakayama, K., Yamashiro, T. & Ohara, N. (2014). Development of a novel plasmid vector pTIO-1 adapted for electrotransformation of *Porphyromonas gingivalis*. *Journal of Microbiological Methods*, 105: 174–179.

[231] Jackson, C. A., Hoffmann, B., Slakeski, N., Cleal, S., Hendtlass, A. J. & Reynolds, E. C. (2000). A consensus *Porphyromonas gingivalis* promoter sequence. *FEMS Microbiology Letters*, 186(1): 133–138.

[232] Chaturvedi, D. & Mahalakshmi, R. (2017). Transmembrane β-barrels: Evolution, folding and energetics. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1859(12): 2467–2482.

[233] Mergulhão, F. J. M., Summers, D. K. & Monteiro, G. A. (2005). Recombinant protein secretion in *Escherichia coli*. *Biotechnology Advances*, 23(3): 177–202.

[234] Schmidt, T. G. & Skerra, A. (2007). The Strep -tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature Protocols*, 2(6): 1528–1535.

[235] Skerra, A. (1994). Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene*, 151(1): 131–135.

[236] Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K. J., Slotboom, D. J., Persson, J. O. & de Gier, J.-W. (2008). Tuning *Escherichia coli* for membrane protein overexpression. *Proceedings of the National Academy of Sciences*, 105(38): 14,371–14,376.

[237] Miroux, B. & Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, 260(3): 289–298.

[238] Joseph, B. C., Pichaimuthu, S. & Srimeenakshi, S. (2015). An overview of the parameters for recombinant protein expression in *Escherichia coli*. *Journal of Cell Science & Therapy*, 06(05).

[239] Loll, P. J. (2014). Membrane proteins, detergents and crystals: What is the state of the art? *Acta Crystallographica. Section F, Structural Biology Communications*, 70(Pt 12): 1576–1583.

[240] Gulati, S., Jamshad, M., Knowles, T. J., Morrison, K. A., Downing, R., Cant, N., Collins, R., Koenderink, J. B., Ford, R. C., Overduin, M., Kerr, I. D., Dafforn, T. R. & Rothnie, A. J. (2014). Detergent-free purification of ABC (ATP-binding-cassette) transporters. *The Biochemical Journal*, 461(2): 269–278.

[241] Dörr, J. M., Koorengevel, M. C., Schäfer, M., Prokofyev, A. V., Scheidelaar, S., van der Cruijsen, E. A. W., Dafforn, T. R., Baldus, M. & Killian, J. A. (2014). Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K+ channel: The power of native nanodiscs. *Proceedings of the National Academy of Sciences*, 111(52): 18,607–18,612.

[242] Jamshad, M., Charlton, J., Lin, Y.-P., Routledge, S. J., Bawa, Z., Knowles, T. J., Overduin, M., Dekker, N., Dafforn, T. R., Bill, R. M., Poyner, D. R. & Wheatley, M. (2015). G-protein coupled receptor solubilization and purification for biophysical analysis and functional studies, in the total absence of detergent. *Bioscience Reports*, 35(2).

[243] Eng, J. K., McCormack, A. L. & Yates, J. R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 5(11): 976–989.

[244] Piggot, T. J., Holdbrook, D. A. & Khalid, S. (2013). Conformational dynamics and membrane interactions of the *E. coli* outer membrane protein FecA: A molecular dynamics simulation study. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1828(2): 284–293.

[245] Hicks, R. (2017). Preparation of membrane models of Gram-negative bacteria and their interaction with antimicrobial peptides studied by CD and NMR. In Hansen, P. R. (editor), *Antimicrobial Peptides: Methods and Protocols*, Methods in Molecular Biology, pp. 231–245. Springer, New York, NY.

[246] Boags, A., Hsu, P.-C., Samsudin, F., Bond, P. J. & Khalid, S. (2017). Progress in molecular dynamics simulations of Gram-negative bacterial cell envelopes. *The Journal of Physical Chemistry Letters*, 8(11): 2513–2518.

[247] Koukos, P. I., Faro, I., van Noort, C. W. & Bonvin, A. M. J. J. (2018). A Membrane Protein Complex Docking Benchmark. *Journal of Molecular Biology*, 430(24): 5246–5256.

[248] Vangone, A., Spinelli, R., Scarano, V., Cavallo, L. & Oliva, R. (2011). COCOMAPS: A web application to analyze and visualize contacts at the interface of biomolecular complexes. *Bioinformatics*, 27(20): 2915–2916.

[249] Chermak, E., Petta, A., Serra, L., Vangone, A., Scarano, V., Cavallo, L. & Oliva, R. (2015). CONSRANK: A server for the analysis, comparison and ranking of docking models based on inter-residue contacts. *Bioinformatics*, 31(9): 1481–1483.

[250] Niedzialkowska, E., Gasiorowska, O., Handing, K. B., Majorek, K. A., Porebski, P. J., Shabalin, I. G., Zasadzinska, E., Cymborowski, M. & Minor, W. (2016). Protein purification and crystallization artifacts: The tale usually not told. *Protein Science*, 25(3): 720–733.

[251] van den Berg, B., Black, P. N., Clemons, W. M. & Rapoport, T. A. (2004). Crystal structure of the long-chain fatty acid transporter FadL. *Science*, 304(5676): 1506–1509.

[252] Kulkarni, S. S., Zhu, Y., Brendel, C. J. & McBride, M. J. (2017). Diverse C-terminal sequences involved in *Flavobacterium johnsoniae* protein secretion. *Journal of Bacteriology*, 199(12).

Appendix A

Strains, Plasmids and Oligonucleotides

Table A.0.1: List of bacterial strains used.

Strain	Description	Antibiotic resistance*	Source
***Escherichia coli* K-12**			
NEB5α	E. coli strain used for molecular cloning and plasmid preparation.	-	NEB
BL21 (DE3)	E. coli strain used recombinant protein expression.	-	NEB
Lemo21 (DE3)	E. coli strain used recombinant protein expression.	Cam	NEB
C41 (DE3)	E. coli strain used recombinant protein expression.	-	Sigma-Aldrich
Porphyromonas gingivalis			
W50	Wild type strain	-	Prof. Mike Curtis
0027	ΔPorV (W50 derived)	Erm	Prof. Mike Curtis
D7	ΔRgpB (W50 derived)	Erm	Prof. Mike Curtis

*Cam - Chloramphenicol; Erm - Erythromycin

Table A.0.2: List of plasmids used for molecular cloning.

Plasmid	Description	Antibiotic resistance*	Source
Expression vectors			
pET-46 EK/LIC	Expression vector with N- His-tag and EK cleavage site	Amp	Thermo-Fisher
pET-46-RgpB-CTD	RgpB-CTD with Sec peptide and N-terminal His-tag and EK cleavage site	Amp	This study
pET-28(b)	Protein expression vector with N- or C- His-tag and EK cleavage site	Kan	Thermo-Fisher
pET-28-OSS-PorV-His	RgpB-CTD with OmpA Sec peptide, C-terminal His-tag and EK cleavage site	Kan	This study
pET-28-OSS-His-PorV	RgpB-CTD with OmpA Sec peptide, N-terminal His-tag and EK cleavage site	Kan	This study
pASK-IBA2C	Protein expression vector with C- Strep-tag and OmpA promoter	Cam	IBA Scientific
pASK-PorV	PorV with OmpA Sec peptide and C-terminal strep-tag	Cam	This study
Shuttle vectors			
PTIO-AI	Shuttle vector used for protein expression in *P. gingivalis*	Amp/ Erm	Dr. Mark Roberts
pTIO-AI-SCH	RgpB-CTD with Sec peptide and N-terminal His-tag	Amp/ Erm	This Study
pTIO-AI-SHC	RgpB-CTD with Sec peptide and C-terminal His-tag	Amp/ Erm	This Study
pTIO-AI-CH	RgpB-CTD with C-terminal His-tag	Amp/ Erm	This Study

*Amp - Ampicillin; Cam - Chloramphenicol; Erm - Erythromycin; Kan - Kanamycin

Table A.0.3: List of oligonucleotides used used for molecular cloning.

Oligo name	Vector	Sequence (5'-3')
RCTD-F	pET46	GACGACGACAAGATGGGTACATCTATTGCCGACGTAG
RCTD-R	pET46	GAGGAGAAGCCCGGTTACTTCACTATAACCTTTTCTGTATACGTC
SHC-Sec-F	pTIO-AI	CATCAAAGGTACCACCATGGATGAAAAAGAATTTTAGCAGGATCGTTTCGATCGTAGC
SHC-Sec-R	pTIO-AI	GGTGGTGGTGGTGGTGTTGTGGGTTGCGACCGCG
SHC-CTD-F	pTIO-AI	CCACAACACCACCACCACCACCACGGTACATCTATTGCCGAC
SHC-CTD-R	pTIO-AI	GGTGATGGTGATGAGATCTGGATCCTTACTTCACTATAACCTTTTCTG
SCH-Sec-F	pTIO-AI	AATTTCATCAAAGGTACCACCATGGATGAAAAAGAATTTTAGCAGGATCGTTTCGATCG
SCH-Sec-R	pTIO-AI	TAGATGTACCTTGTGGGTTGCGACCGCG
SCH-CTD-F	pTIO-AI	GGTCGCAACCCACAAGGTACATCTATTGCCGACG
SCH-CTD-R	pTIO-AI	GGTGATGGTGATGAGATCTGGATCCCTTCACTATAACCTTTTCTG
CH-CTD-F	pTIO-AI	CTAATTTCATCAAAGGTACCACCATGGGTACATCTATTGCCGACG
PorV-ASK_-F	pASK-IBA2C	CGAGGTACCGCTCAGGAGCAACTGAATGTG
PorV-ASK-R	pASK-IBA2C	GGTGCTCGAGGTGGAACAAATTGCGCAATCC
PorV-28-F	pET28-OSS	GACAAGCTTGCTCAGGAGCAACTGAATG
PorV-28-R	pET28-OSS	CCAGCGGCCGCTCAGTGGAACAAATTGCGCAATCC
PorV-28(His)-R	pET28-OSS	CCACTCGAGGTGGAACAAATTGCGCAATCC

Appendix B

RgpB-CTD NMR Assignments

Number	Shift	SD	Assign Name	Residue
1	55.86958	0.02816	Ca	654His
2	174.42915	0	C	654His
3	29.68379	0.03785	Cb	654His
4	4.66464	0	Ha	654His
5	3.18144	0	Hba	654His
6	3.18144	0	Hbb	654His
7	8.10138	0.00302	H	655Val
8	121.58626	0.03530	N	655Val
9	62.28577	0.03112	Ca	655Val
10	32.85398	0.04235	Cb	655Val
11	175.55527	0.00730	C	655Val
12	21.00443	0	Cgb	655Val
13	20.43368	0	Cga	655Val
14	8.37434	0.00594	H	656Asp
15	122.95294	0.02801	N	656Asp
16	41.29520	0.05142	Cb	656Asp
17	54.60766	0.03602	Ca	656Asp
18	176.05622	0.03887	C	656Asp
19	8.15443	0.00211	H	657Asp
20	120.23387	0.05826	N	657Asp
21	55.00993	0	Ca	657Asp
22	41.59007	0	Cb	657Asp
23	176.26944	0	C	657Asp
24	41.03181	0.10588	Cb	658Asp
25	176.28368	0.01278	C	658Asp

Number	Shift	SD	Assign Name	Residue
26	54.67223	0.03679	Ca	658Asp
27	8.21067	0.00317	H	659Asp
28	120.49930	0.11444	N	659Asp
29	40.91372	0.08361	Cb	659Asp
30	54.67539	0.08397	Ca	659Asp
31	176.63999	0.00890	C	659Asp
32	4.57781	0.00651	Ha	659Asp
33	2.72897	0	Hba	659Asp
34	2.72897	0	Hbb	659Asp
35	8.11428	0.00598	H	660Lys
36	120.50508	0.03479	N	660Lys
37	56.53956	0.08878	Ca	660Lys
38	32.49050	0.04465	Cb	660Lys
39	177.06874	0.01493	C	660Lys
40	28.79369	0	Cd	660Lys
41	24.68451	0	Cg	660Lys
42	40.92490	0	Ce	660Lys
43	4.28584	0.00813	Ha	660Lys
44	3.01578	0.00805	Hea	660Lys
45	3.01984	0.01211	Heb	660Lys
46	1.88717	0.00347	Hbb	660Lys
47	1.81472	0.00363	Hba	660Lys
48	1.45712	8.34826e-04	Hgb	660Lys
49	1.42251	0.00689	Hga	660Lys
50	1.67298	2.74782e-04	Hda	660Lys
51	1.67298	2.74782e-04	Hdb	660Lys
52	8.22061	0.00495	H	661Met
53	119.64824	0.02854	N	661Met
54	55.69198	0.05325	Ca	661Met
55	176.96730	0.01092	C	661Met
56	32.40414	0.08732	Cb	661Met
57	31.92803	0.07354	Cg	661Met
58	4.47216	0.00491	Ha	661Met
59	2.62992	0.00279	Hgb	661Met
60	2.53671	0.00438	Hga	661Met
61	2.11579	0.00783	Hbb	661Met
62	2.07549	0.00541	Hba	661Met
63	8.30490	0.00708	H	662Gly
64	109.52039	0.02917	N	662Gly
65	45.62336	0.05679	Ca	662Gly
66	174.35288	0.01662	C	662Gly
67	4.00637	0.00570	Haa	662Gly

Number	Shift	SD	Assign Name	Residue
68	4.05276	0.00887	Hab	662Gly
69	7.98536	0.00853	H	663Thr
70	113.14732	0.04785	N	663Thr
71	61.79305	0.06530	Ca	663Thr
72	69.80660	0.05454	Cb	663Thr
73	174.65896	0.04811	C	663Thr
74	21.50787	0.04413	Cg2	663Thr
75	1.20312	0.00638	Hg2*	663Thr
76	4.39981	0.01413	Ha	663Thr
77	4.24995	0.00838	Hb	663Thr
78	8.33806	0.00670	H	664Ser
79	118.40031	0.12178	N	664Ser
80	58.19256	0.04653	Ca	664Ser
81	63.89416	0.10650	Cb	664Ser
82	174.44497	0.04236	C	664Ser
83	3.87665	0.00113	Hba	664Ser
84	3.87665	0.00113	Hbb	664Ser
85	4.53928	0.00297	Ha	664Ser
86	8.09999	0.00762	H	665Ile
87	122.33016	0.07938	N	665Ile
88	61.24325	0.07955	Ca	665Ile
89	38.69700	0.09746	Cb	665Ile
90	175.83455	0.01232	C	665Ile
91	27.27774	0.04687	Cg1	665Ile
92	17.62296	0.05758	Cg2	665Ile
93	13.10261	0.03864	Cd1	665Ile
94	4.19425	0.00359	Ha	665Ile
95	1.88170	0.00437	Hb	665Ile
96	1.44509	0.02256	Hg1b	665Ile
97	1.18340	0.00769	Hg1a	665Ile
98	0.91263	0.01745	Hg2*	665Ile
99	0.86060	0.00780	Hd1*	665Ile
100	8.17938	0.00744	H	666Ala
101	127.11133	0.04919	N	666Ala
102	52.49050	0.10132	Ca	666Ala
103	19.49263	0.01975	Cb	666Ala
104	177.25588	0.00458	C	666Ala
105	4.38359	0.02194	Ha	666Ala
106	1.39290	0.02431	Hb*	666Ala
107	8.15461	0.00574	H	667Asp
108	119.80123	0.08065	N	667Asp
109	41.27910	0.03711	Cb	667Asp

Number	Shift	SD	Assign Name	Residue
110	54.47759	0.06585	Ca	667Asp
111	176.30087	0.01234	C	667Asp
112	4.65004	0.01062	Ha	667Asp
113	2.61880	0.00932	Hba	667Asp
114	2.71782	0.01186	Hbb	667Asp
115	7.90634	0.00872	H	668Val
116	119.51860	0.02170	N	668Val
117	175.92552	0.00982	C	668Val
118	62.22452	0.05987	Ca	668Val
119	32.65996	0.04874	Cb	668Val
120	20.52764	0.00416	Cga	668Val
121	4.12419	0.00913	Ha	668Val
122	2.13155	0.00659	Hb	668Val
123	0.93748	0.00707	Hga*	668Val
124	0.93885	0.01017	Hgb*	668Val
125	21.05940	0	Cgb	668Val
126	8.24986	0.00656	H	669Ala
127	126.59109	0.04365	N	669Ala
128	52.58668	0.07991	Ca	669Ala
129	19.29508	0.02201	Cb	669Ala
130	177.41375	0.00242	C	669Ala
131	4.32394	0.00675	Ha	669Ala
132	1.39407	0.00662	Hb*	669Ala
133	8.19675	0.00531	H	670Asn
134	117.76598	0.02671	N	670Asn
135	174.76373	0.00242	C	670Asn
136	38.96364	0.08587	Cb	670Asn
137	53.17687	0.09347	Ca	670Asn
138	4.70058	0.00584	Ha	670Asn
139	2.72817	0.00768	Hba	670Asn
140	2.83228	0.00330	Hbb	670Asn
141	7.57286	0.00451	Hd2b	670Asn
142	112.61077	0.04558	Nd2	670Asn
143	6.80883	0.00799	Hd2a	670Asn
144	8.20401	0.00684	H	671Asp
145	120.28347	0.05187	N	671Asp
146	175.52914	5.04637e-06	C	671Asp
147	54.42832	0.08053	Ca	671Asp
148	40.99764	0.04319	Cb	671Asp
149	4.57469	0.01142	Ha	671Asp
150	2.63892	0.00673	Hba	671Asp
151	2.68397	0.02363	Hbb	671Asp

Number	Shift	SD	Assign Name	Residue
152	7.93795	0.00848	H	672Lys
153	120.54407	0.04445	N	672Lys
154	32.89549	0.03932	Cb	672Lys
155	53.56980	0.03432	Ca	672Lys
156	175.28428	0	C	672Lys
157	42.03973	0.00795	Ce	672Lys
158	29.00793	0.01901	Cd	672Lys
159	0.98228	0.00499	Hba	672Lys
160	1.35758	0.00574	Hbb	672Lys
161	4.26321	0.00381	Ha	672Lys
162	24.72070	0.02308	Cg	672Lys
163	2.94086	0.00761	Hea	672Lys
164	2.94086	0.00761	Heb	672Lys
165	1.51348	0.00784	Hdb	672Lys
166	1.46637	0.01743	Hda	672Lys
167	1.13348	0.01143	Hgb	672Lys
168	1.09120	0.00651	Hga	672Lys
169	63.21311	0.03499	Ca	673Pro
170	174.25640	0.00106	C	673Pro
171	29.49247	0.06029	Cb	673Pro
172	49.84446	0.05109	Cd	673Pro
173	27.29547	0.07085	Cg	673Pro
174	4.64992	0.00542	Ha	673Pro
175	3.54160	0.00426	Hdb	673Pro
176	2.91607	0.00404	Hda	673Pro
177	2.33659	0.00438	Hbb	673Pro
178	1.97345	0.00463	Hba	673Pro
179	2.15235	0.00490	Hgb	673Pro
180	1.55733	0.00514	Hga	673Pro
181	7.11406	0.00699	H	674Tyr
182	115.51116	0.01207	N	674Tyr
183	40.63953	0.06619	Cb	674Tyr
184	55.46252	0.07901	Ca	674Tyr
185	173.25380	0.03734	C	674Tyr
186	5.30269	0.00685	Ha	674Tyr
187	2.90899	0.00474	Hbb	674Tyr
188	2.74358	0.00423	Hba	674Tyr
189	6.19310	0.01619	He*	674Tyr
190	6.21451	0.01557	Hd*	674Tyr
191	117.76262	0	Ce*	674Tyr
192	133.10071	0	Cd*	674Tyr
193	8.85859	0.00771	H	675Thr

Number	Shift	SD	Assign Name	Residue
194	111.82832	0.02736	N	675Thr
195	61.15733	0.06654	Ca	675Thr
196	71.84521	0.05852	Cb	675Thr
197	173.18849	0.00846	C	675Thr
198	21.48586	0.06440	Cg2	675Thr
199	1.14010	0.00618	Hg2*	675Thr
200	4.03149	0.00630	Hb	675Thr
201	4.55892	0.00518	Ha	675Thr
202	8.55328	0.01004	H	676Val
203	123.85466	0.02347	N	676Val
204	34.41013	0.08954	Cb	676Val
205	61.16906	0.07664	Ca	676Val
206	173.93550	0.00925	C	676Val
207	21.60773	0	Cgb	676Val
208	4.81053	0.00877	Ha	676Val
209	2.00404	0.00341	Hb	676Val
210	0.90674	0.00420	Hgb*	676Val
211	0.80614	0.00516	Hga*	676Val
212	21.57452	0.07996	Cga	676Val
213	9.08063	0.00782	H	677Ala
214	130.45282	0.02046	N	677Ala
215	50.98016	0.04576	Ca	677Ala
216	22.42653	0.06090	Cb	677Ala
217	175.15918	0.02061	C	677Ala
218	4.79431	0.00347	Ha	677Ala
219	1.33128	0.00849	Hb*	677Ala
220	8.47949	0.00880	H	678Val
221	121.60707	0.06826	N	678Val
222	33.85632	0.05812	Cb	678Val
223	60.86556	0.07876	Ca	678Val
224	175.85541	0.00384	C	678Val
225	4.85317	0.00591	Ha	678Val
226	1.94030	0.00571	Hb	678Val
227	0.86460	0.01266	Hga*	678Val
228	0.86637	0.01285	Hgb*	678Val
229	21.78443	0	Cgb	678Val
230	21.32922	0	Cga	678Val
231	8.51686	0.00828	H	679Ser
232	123.04399	0.03027	N	679Ser
233	56.05662	0.09199	Ca	679Ser
234	63.47448	0.08779	Cb	679Ser
235	175.25351	0.00637	C	679Ser

Number	Shift	SD	Assign Name	Residue
236	3.65032	0.00546	Hba	679Ser
237	3.77133	0.00623	Hbb	679Ser
238	4.69269	0.00888	Ha	679Ser
239	9.11266	0.00950	H	680Gly
240	119.04147	0.02002	N	680Gly
241	173.69744	0.00959	C	680Gly
242	47.43756	0.07301	Ca	680Gly
243	4.01640	0.00713	Hab	680Gly
244	3.66738	0.00735	Haa	680Gly
245	8.58975	0.00994	H	681Lys
246	126.05018	0.01893	N	681Lys
247	55.09503	0.05616	Ca	681Lys
248	31.73769	0.03908	Cb	681Lys
249	174.39004	0.01354	C	681Lys
250	28.58934	0.08728	Cd	681Lys
251	25.15078	0.05934	Cg	681Lys
252	42.52459	0.10685	Ce	681Lys
253	4.33939	0.00519	Ha	681Lys
254	2.93224	0.00784	Heb	681Lys
255	2.93113	0.00422	Hea	681Lys
256	1.58867	0.00738	Hba	681Lys
257	2.28413	0.00417	Hbb	681Lys
258	1.51746	0.00239	Hdb	681Lys
259	1.23995	0.00459	Hga	681Lys
260	1.36892	0.00569	Hgb	681Lys
261	1.51746	0.00261	Hda	681Lys
262	8.10571	0.00698	H	682Thr
263	115.38384	0.02616	N	682Thr
264	62.42962	0.05486	Ca	682Thr
265	69.56692	0.07309	Cb	682Thr
266	174.39710	0.06215	C	682Thr
267	22.69044	0.05105	Cg2	682Thr
268	0.97368	0.00520	Hg2*	682Thr
269	4.10477	0.00399	Hb	682Thr
270	4.64283	0.01110	Ha	682Thr
271	9.37770	0.00980	H	683Ile
272	129.64137	0.03717	N	683Ile
273	40.18077	0.05879	Cb	683Ile
274	61.01442	0.10307	Ca	683Ile
275	174.90204	0.00761	C	683Ile
276	27.00995	0.05892	Cg1	683Ile
277	16.93583	0.04533	Cg2	683Ile

Number	Shift	SD	Assign Name	Residue
278	13.67812	0.03513	Cd1	683Ile
279	4.54742	0.00368	Ha	683Ile
280	1.77582	0.00473	Hb	683Ile
281	0.66493	0.01176	Hg1a	683Ile
282	0.64844	0.00814	Hg2*	683Ile
283	0.40133	0.00542	Hd1*	683Ile
284	1.60034	0.00245	Hg1b	683Ile
285	9.08959	0.00907	H	684Thr
286	123.39000	0.05619	N	684Thr
287	61.56314	0.09346	Ca	684Thr
288	70.06474	0.04711	Cb	684Thr
289	174.21129	0.03853	C	684Thr
290	20.81430	0.03305	Cg2	684Thr
291	1.11294	0.00494	Hg2*	684Thr
292	3.91526	0.00623	Hb	684Thr
293	5.32527	0.00707	Ha	684Thr
294	9.01814	0.00968	H	685Val
295	126.74109	0.04321	N	685Val
296	59.74014	0.05804	Ca	685Val
297	34.74174	0.08488	Cb	685Val
298	173.96552	0.02706	C	685Val
299	22.03941	0.01159	Cgb	685Val
300	5.09763	0.00453	Ha	685Val
301	1.81044	0.00971	Hb	685Val
302	0.82126	0.00449	Hgb*	685Val
303	0.81420	0.00257	Hga*	685Val
304	21.42157	0.00498	Cga	685Val
305	8.56339	0.00789	H	686Glu
306	124.83958	0.03950	N	686Glu
307	175.52219	0.00572	C	686Glu
308	33.41255	0.05881	Cb	686Glu
309	55.32158	0.08222	Ca	686Glu
310	36.43063	0.03185	Cg	686Glu
311	2.12197	0.02103	Hga	686Glu
312	2.20617	0.00498	Hgb	686Glu
313	1.97557	0.01092	Hbb	686Glu
314	1.93146	0.01147	Hba	686Glu
315	5.03525	0.00675	Ha	686Glu
316	8.60438	0.00756	H	687Ser
317	117.47216	0.02992	N	687Ser
318	171.64466	0	C	687Ser
319	56.44729	0.05121	Ca	687Ser

Number	Shift	SD	Assign Name	Residue
320	64.37910	0.03811	Cb	687Ser
321	3.85476	0.00388	Hbb	687Ser
322	4.79784	0.00759	Ha	687Ser
323	3.50274	0.00463	Hba	687Ser
324	177.22470	0.01716	C	688Pro
325	64.92859	0.09364	Ca	688Pro
326	31.94094	0.06584	Cb	688Pro
327	50.97306	0.01910	Cd	688Pro
328	27.53251	0.07509	Cg	688Pro
329	4.27970	0.00502	Ha	688Pro
330	3.81764	0.00753	Hdb	688Pro
331	3.74946	0.00572	Hda	688Pro
332	2.13921	0.00532	Hgb	688Pro
333	1.99635	0.00918	Hga	688Pro
334	1.97637	0.00747	Hba	688Pro
335	2.31154	0.00488	Hbb	688Pro
336	8.54346	0.00888	H	689Ala
337	121.96132	0.02936	N	689Ala
338	176.03788	0.00793	C	689Ala
339	17.84220	0.08554	Cb	689Ala
340	53.13325	0.04418	Ca	689Ala
341	4.24497	0.00890	Ha	689Ala
342	1.51589	0.00970	Hb*	689Ala
343	8.16685	0.00515	H	690Ala
344	121.73424	0.01856	N	690Ala
345	177.45097	0.00341	C	690Ala
346	51.64229	0.04747	Ca	690Ala
347	22.60003	0.03554	Cb	690Ala
348	4.87615	0.00586	Ha	690Ala
349	1.27984	0.00452	Hb*	690Ala
350	8.65315	0.00624	H	691Gly
351	107.42327	0.02940	N	691Gly
352	45.49228	0.06996	Ca	691Gly
353	172.99354	0.00958	C	691Gly
354	4.42846	0.00439	Hab	691Gly
355	3.79564	0.00666	Haa	691Gly
356	8.79797	0.00711	H	692Leu
357	126.41654	0.02158	N	692Leu
358	53.27044	0.06852	Ca	692Leu
359	176.13910	0.00786	C	692Leu
360	44.59544	0.06861	Cb	692Leu
361	26.32392	0.09939	Cg	692Leu

Number	Shift	SD	Assign Name	Residue
362	26.03448	0.05406	Cda	692Leu
363	5.35121	0.00504	Ha	692Leu
364	1.72425	0.00617	Hbb	692Leu
365	0.95594	0.00895	Hba	692Leu
366	1.55558	0.00309	Hg	692Leu
367	0.69444	0.00562	Hdb*	692Leu
368	0.52921	0.00553	Hda*	692Leu
369	25.75127	0.04506	Cdb	692Leu
370	8.66285	0.00826	H	693Thr
371	117.64714	0.02762	N	693Thr
372	72.14565	0.08432	Cb	693Thr
373	61.52367	0.08683	Ca	693Thr
374	172.07313	0.01491	C	693Thr
375	23.13174	0.04234	Cg2	693Thr
376	0.69394	0.00534	Hg2*	693Thr
377	3.67871	0.00590	Hb	693Thr
378	4.88727	0.00673	Ha	693Thr
379	8.15453	0.00562	H	694Ile
380	123.26075	0.03371	N	694Ile
381	60.02762	0.06473	Ca	694Ile
382	39.81869	0.04329	Cb	694Ile
383	173.82816	0.00894	C	694Ile
384	28.01645	0.02139	Cg1	694Ile
385	18.16016	0.03739	Cg2	694Ile
386	13.45886	0.06053	Cd1	694Ile
387	5.02119	0.00528	Ha	694Ile
388	1.01546	0.00692	Hb	694Ile
389	1.28176	0.00740	Hg1b	694Ile
390	0.77090	0.00346	Hg1a	694Ile
391	0.34861	0.00610	Hd1*	694Ile
392	0.49657	0.00873	Hg2*	694Ile
393	9.49410	0.00878	H	695Phe
394	126.51016	0.02945	N	695Phe
395	42.91501	0.03589	Cb	695Phe
396	55.63028	0.04502	Ca	695Phe
397	176.16901	0.00960	C	695Phe
398	5.64905	0.01138	Ha	695Phe
399	2.95218	0.01417	Hbb	695Phe
400	2.93836	0.01010	Hba	695Phe
401	7.19841	0.01087	He*	695Phe
402	6.98724	0.00668	Hd*	695Phe
403	131.86651	0	Cd*	695Phe

Number	Shift	SD	Assign Name	Residue
404	131.06837	0	Ce*	695Phe
405	8.80330	0.00655	H	696Asp
406	120.18541	0.02724	N	696Asp
407	52.14285	0.07350	Ca	696Asp
408	41.71768	0.06672	Cb	696Asp
409	178.62854	0.01096	C	696Asp
410	4.76942	0.00692	Ha	696Asp
411	3.39505	0.01055	Hbb	696Asp
412	2.73314	0.00653	Hba	696Asp
413	8.31211	0.00685	H	697Met
414	116.49600	0.01965	N	697Met
415	57.63792	0.07556	Ca	697Met
416	176.62508	0.01733	C	697Met
417	31.78302	0.08340	Cb	697Met
418	31.85671	9.84622e-04	Cg	697Met
419	4.25394	0.00461	Ha	697Met
420	2.53965	0.00516	Hgb	697Met
421	2.53946	0.00544	Hga	697Met
422	2.14776	0.00279	Hba	697Met
423	2.14776	0.00279	Hbb	697Met
424	8.21672	0.00522	H	698Asn
425	117.76625	0.01997	N	698Asn
426	175.50540	0.01640	C	698Asn
427	52.96821	0.05628	Ca	698Asn
428	39.36413	0.05660	Cb	698Asn
429	4.90786	0.00708	Ha	698Asn
430	2.99988	0.00888	Hbb	698Asn
431	2.80587	0.00974	Hba	698Asn
432	6.93906	0.00349	Hd2a	698Asn
433	113.99694	0.05897	Nd2	698Asn
434	7.69215	0.00508	Hd2b	698Asn
435	8.26639	0.00530	H	699Gly
436	108.51929	0.01825	N	699Gly
437	174.01257	0.00214	C	699Gly
438	45.55177	0.08147	Ca	699Gly
439	4.18232	0.00809	Hab	699Gly
440	3.62057	0.00577	Haa	699Gly
441	8.45750	0.00592	H	700Arg
442	122.07652	0.06898	N	700Arg
443	176.31053	0.04752	C	700Arg
444	55.74249	0.09537	Ca	700Arg
445	30.52252	0.02846	Cb	700Arg

Number	Shift	SD	Assign Name	Residue
446	43.39462	0.06462	Cd	700Arg
447	27.64535	0.06887	Cg	700Arg
448	4.39778	0.00527	Ha	700Arg
449	3.25177	0.00648	Hdb	700Arg
450	3.25143	0.00574	Hda	700Arg
451	1.65171	0.00497	Hgb	700Arg
452	1.52540	0.00497	Hga	700Arg
453	1.85471	0.00789	Hba	700Arg
454	1.88054	0.01580	Hbb	700Arg
455	8.77273	0.00981	H	701Arg
456	125.65842	0.02827	N	701Arg
457	56.54648	0.07843	Ca	701Arg
458	30.16337	0.03003	Cb	701Arg
459	176.46785	0.02106	C	701Arg
460	43.62962	0.06336	Cd	701Arg
461	26.96914	0.08197	Cg	701Arg
462	3.97677	0.00477	Ha	701Arg
463	3.03552	0.00729	Hdb	701Arg
464	3.03435	0.00666	Hda	701Arg
465	1.62128	0.00420	Hba	701Arg
466	1.72496	0.00428	Hbb	701Arg
467	1.12624	0.00236	Hgb	701Arg
468	1.00957	0.00360	Hga	701Arg
469	8.68655	0.00715	H	702Val
470	121.70460	0.02362	N	702Val
471	32.80365	0.04604	Cb	702Val
472	61.70564	0.06525	Ca	702Val
473	175.38832	0.01873	C	702Val
474	22.64053	0.05254	Cgb	702Val
475	20.16865	0.06032	Cga	702Val
476	4.43197	0.00694	Ha	702Val
477	2.28484	0.00639	Hb	702Val
478	0.59054	0.00525	Hga*	702Val
479	1.01748	0.00473	Hgb*	702Val
480	7.12991	0.00617	H	703Ala
481	120.33529	0.01172	N	703Ala
482	22.54005	0.03414	Cb	703Ala
483	52.79636	0.05409	Ca	703Ala
484	175.38992	0.06522	C	703Ala
485	4.47635	0.00738	Ha	703Ala
486	1.02148	0.00837	Hb*	703Ala
487	8.21037	0.00536	H	704Thr

Number	Shift	SD	Assign Name	Residue
488	113.82951	0.02992	N	704Thr
489	60.30833	0.06525	Ca	704Thr
490	71.85718	0.07424	Cb	704Thr
491	172.38198	0.00143	C	704Thr
492	20.74984	0.02872	Cg2	704Thr
493	1.12145	0.00649	Hg2*	704Thr
494	3.98407	0.00723	Hb	704Thr
495	5.13329	0.00759	Ha	704Thr
496	8.96717	0.00711	H	705Ala
497	125.04904	0.03369	N	705Ala
498	175.28830	0.02387	C	705Ala
499	52.07666	0.03113	Ca	705Ala
500	23.72654	0.06934	Cb	705Ala
501	4.65333	0.00583	Ha	705Ala
502	1.54816	0.00747	Hb*	705Ala
503	8.41444	0.00902	H	706Lys
504	119.39519	0.01909	N	706Lys
505	55.34111	0.09643	Ca	706Lys
506	34.56470	0.03127	Cb	706Lys
507	178.26150	0.00611	C	706Lys
508	41.96423	0.08247	Ce	706Lys
509	29.02724	0.06677	Cd	706Lys
510	24.88718	0.04281	Cg	706Lys
511	5.08356	0.00645	Ha	706Lys
512	2.94478	0.01003	Hea	706Lys
513	2.94520	0.01141	Heb	706Lys
514	1.66208	0.00931	Hda	706Lys
515	1.73983	0.02088	Hba	706Lys
516	1.74642	0.00681	Hbb	706Lys
517	1.45422	0.00817	Hgb	706Lys
518	1.38869	0.00297	Hga	706Lys
519	1.66609	0.00802	Hdb	706Lys
520	8.90320	0.00771	H	707Asn
521	116.94023	0.02485	N	707Asn
522	58.55316	0.08741	Ca	707Asn
523	38.00781	0.04226	Cb	707Asn
524	175.04104	0.01896	C	707Asn
525	4.62861	0.00538	Ha	707Asn
526	3.51210	0.00822	Hbb	707Asn
527	2.92509	0.00303	Hba	707Asn
528	113.59247	0.05335	Nd2	707Asn
529	7.59046	0.00349	Hd2b	707Asn

Number	Shift	SD	Assign Name	Residue
530	6.88157	0.00353	Hd2a	707Asn
531	9.27432	0.00863	H	708Arg
532	124.45264	0.03226	N	708Arg
533	32.84200	0.05644	Cb	708Arg
534	55.61096	0.08659	Ca	708Arg
535	173.38366	0.00679	C	708Arg
536	43.25407	0.09377	Cd	708Arg
537	27.73047	0.03819	Cg	708Arg
538	5.02953	0.00742	Ha	708Arg
539	3.15278	0.00713	Hda	708Arg
540	3.15316	0.00637	Hdb	708Arg
541	1.49798	0.00586	Hgb	708Arg
542	1.41803	0.00365	Hga	708Arg
543	1.75654	0.01742	Hba	708Arg
544	1.77474	0.01812	Hbb	708Arg
545	8.94616	0.00774	H	709Met
546	123.41306	0.01204	N	709Met
547	54.43430	0.08829	Ca	709Met
548	37.96128	0.04012	Cb	709Met
549	172.73536	0.00526	C	709Met
550	30.37447	0.06792	Cg	709Met
551	4.63077	0.00534	Ha	709Met
552	1.76235	0.01628	Hba	709Met
553	2.22142	0.00456	Hgb	709Met
554	2.02708	0.00457	Hga	709Met
555	1.81191	0.01172	Hbb	709Met
556	8.28712	0.00786	H	710Val
557	125.50878	0.03978	N	710Val
558	33.40827	0.06098	Cb	710Val
559	61.82893	0.08900	Ca	710Val
560	175.04587	0.00286	C	710Val
561	4.73632	0.00753	Ha	710Val
562	1.90493	0.00728	Hb	710Val
563	21.13487	0.07576	Cga	710Val
564	20.60193	0.05646	Cgb	710Val
565	0.90391	0.01065	Hgb*	710Val
566	0.76581	0.00356	Hga*	710Val
567	9.34520	0.00931	H	711Phe
568	129.26617	0.10846	N	711Phe
569	172.58649	0.00747	C	711Phe
570	56.15282	0.05503	Ca	711Phe
571	42.97462	0.06285	Cb	711Phe

Number	Shift	SD	Assign Name	Residue
572	4.84185	0.00785	Ha	711Phe
573	2.73228	0.00618	Hbb	711Phe
574	2.50128	0.00327	Hba	711Phe
575	6.87682	0.00807	Hd*	711Phe
576	6.92728	0.00182	He*	711Phe
577	131.18264	0	Ce*	711Phe
578	131.17227	0	Cd*	711Phe
579	7.26497	0	Hz	711Phe
580	7.91128	0.00872	H	712Glu
581	129.22518	0.03111	N	712Glu
582	173.16255	0.00499	C	712Glu
583	55.72020	0.09806	Ca	712Glu
584	28.66367	0.05276	Cb	712Glu
585	35.50736	0.04975	Cg	712Glu
586	4.26639	0.00467	Ha	712Glu
587	1.71229	0.00598	Hbb	712Glu
588	1.60542	0.00331	Hba	712Glu
589	1.99384	0.00292	Hga	712Glu
590	1.99384	0.00292	Hgb	712Glu
591	8.46126	0.00872	H	713Ala
592	130.37577	0.02797	N	713Ala
593	22.21586	0.02686	Cb	713Ala
594	50.97904	0.05121	Ca	713Ala
595	176.16221	0.00750	C	713Ala
596	4.05704	0.00406	Ha	713Ala
597	1.20658	0.00374	Hb*	713Ala
598	7.52783	0.00673	H	714Gln
599	115.26354	0.03336	N	714Gln
600	30.79684	0.04810	Cb	714Gln
601	54.14264	0.06578	Ca	714Gln
602	175.50559	0.00409	C	714Gln
603	33.76967	0.06998	Cg	714Gln
604	4.48091	0.00410	Ha	714Gln
605	1.94325	0.00338	Hba	714Gln
606	2.25119	0.00412	Hbb	714Gln
607	2.45719	0.00289	Hga	714Gln
608	2.45741	0.00289	Hgb	714Gln
609	8.32814	0.00643	H	715Asn
610	117.84555	0.06451	N	715Asn
611	175.81601	0.00388	C	715Asn
612	38.13380	0.05908	Cb	715Asn
613	54.65494	0.08375	Ca	715Asn

Number	Shift	SD	Assign Name	Residue
614	2.73165	0.01169	Hbb	715Asn
615	2.72253	0.00804	Hba	715Asn
616	4.42321	0.00568	Ha	715Asn
617	110.58652	0.02409	Nd2	715Asn
618	6.71165	0.00484	Hd2a	715Asn
619	7.48536	0.00408	Hd2b	715Asn
620	8.95097	0.00874	H	716Gly
621	111.10024	0.02000	N	716Gly
622	170.56229	0.01096	C	716Gly
623	44.78995	0.10055	Ca	716Gly
624	4.30883	0.00765	Hab	716Gly
625	3.91274	0.00568	Haa	716Gly
626	7.44187	0.00765	H	717Val
627	119.25553	0.01722	N	717Val
628	33.32585	0.03552	Cb	717Val
629	61.54045	0.10047	Ca	717Val
630	174.89695	0.00994	C	717Val
631	21.21844	0.09458	Cga	717Val
632	4.80449	0.00330	Ha	717Val
633	1.84244	0.00306	Hb	717Val
634	0.89804	0.00847	Hga*	717Val
635	0.96422	0.01137	Hgb*	717Val
636	21.44924	0.08622	Cgb	717Val
637	9.11396	0.00988	H	718Tyr
638	124.48682	0.10374	N	718Tyr
639	56.80569	0.03805	Ca	718Tyr
640	41.87417	0.03143	Cb	718Tyr
641	174.38296	0.00919	C	718Tyr
642	4.93319	0.00792	Ha	718Tyr
643	2.81255	0.00576	Hbb	718Tyr
644	2.45865	0.00578	Hba	718Tyr
645	6.91201	0.00792	Hd*	718Tyr
646	7.02175	0.00733	He*	718Tyr
647	133.76566	0	Cd*	718Tyr
648	117.72826	0	Ce*	718Tyr
649	8.92340	0.00850	H	719Ala
650	123.59545	0.03152	N	719Ala
651	175.69447	0.00287	C	719Ala
652	50.43756	0.07016	Ca	719Ala
653	20.82090	0.05310	Cb	719Ala
654	5.17829	0.00672	Ha	719Ala
655	1.29829	0.00552	Hb*	719Ala

Number	Shift	SD	Assign Name	Residue
656	9.20919	0.00768	H	720Val
657	122.98456	0.03200	N	720Val
658	60.59649	0.06413	Ca	720Val
659	174.05461	0.01052	C	720Val
660	4.70756	0.00895	Ha	720Val
661	1.93511	0.00328	Hb	720Val
662	33.26912	0.04362	Cb	720Val
663	0.70875	0.00247	Hga*	720Val
664	21.58135	0.04118	Cga	720Val
665	20.80102	0.05866	Cgb	720Val
666	0.77819	0.00539	Hgb*	720Val
667	9.42886	0.00949	H	721Arg
668	127.91363	0.02610	N	721Arg
669	54.67555	0.08677	Ca	721Arg
670	175.01951	0.00305	C	721Arg
671	33.16420	0.04740	Cb	721Arg
672	43.79750	0.07211	Cd	721Arg
673	27.51193	0.06468	Cg	721Arg
674	5.30580	0.00571	Ha	721Arg
675	3.16708	0.00603	Hdb	721Arg
676	3.16469	0.00455	Hda	721Arg
677	1.87294	0.00405	Hbb	721Arg
678	1.79556	0.01120	Hba	721Arg
679	1.58754	0.00523	Hgb	721Arg
680	1.48766	0.00506	Hga	721Arg
681	9.20153	0.00906	H	722Ile
682	128.52152	0.02628	N	722Ile
683	40.17126	0.05860	Cb	722Ile
684	61.16824	0.05535	Ca	722Ile
685	173.99450	0.01784	C	722Ile
686	28.75165	0.06815	Cg1	722Ile
687	20.03606	0.05859	Cg2	722Ile
688	14.96618	0.03868	Cd1	722Ile
689	4.47789	0.00384	Ha	722Ile
690	1.56442	0.00536	Hb	722Ile
691	0.90549	0.00538	Hg1b	722Ile
692	0.34759	0.00555	Hg1a	722Ile
693	0.59347	0.00441	Hg2*	722Ile
694	-0.16047	0.00725	Hd1*	722Ile
695	9.43607	0.01042	H	723Ala
696	130.95010	0.02417	N	723Ala
697	176.94903	0.01131	C	723Ala

Number	Shift	SD	Assign Name	Residue
698	51.26885	0.05962	Ca	723Ala
699	19.60115	0.06771	Cb	723Ala
700	5.19097	0.00615	Ha	723Ala
701	1.38127	0.00544	Hb*	723Ala
702	8.28765	0.00559	H	724Thr
703	116.92389	0.03206	N	724Thr
704	173.86177	0.00604	C	724Thr
705	70.32979	0.04582	Cb	724Thr
706	60.12222	0.08918	Ca	724Thr
707	21.40415	0.05495	Cg2	724Thr
708	1.38112	0.00510	Hg2*	724Thr
709	4.56815	0.00436	Hb	724Thr
710	4.93769	0.00777	Ha	724Thr
711	8.88349	0.00944	H	725Glu
712	121.72214	0.01797	N	725Glu
713	28.65289	0.06692	Cb	725Glu
714	58.46822	0.06788	Ca	725Glu
715	176.93753	0.00869	C	725Glu
716	36.16607	0.06145	Cg	725Glu
717	4.08217	0.00569	Ha	725Glu
718	2.37546	0.00476	Hgb	725Glu
719	2.31608	0.01298	Hga	725Glu
720	2.12327	0.01218	Hba	725Glu
721	2.12937	0.01150	Hbb	725Glu
722	8.40897	0.00812	H	726Gly
723	108.02226	0.02333	N	726Gly
724	173.87573	0.00718	C	726Gly
725	45.51506	0.06669	Ca	726Gly
726	3.98456	0.01711	Hab	726Gly
727	3.97405	0.00674	Haa	726Gly
728	7.61497	0.00445	H	727Lys
729	120.51717	0.03703	N	727Lys
730	175.08002	0.01833	C	727Lys
731	55.99346	0.08347	Ca	727Lys
732	35.46609	0.07140	Cb	727Lys
733	24.75020	0.03016	Cg	727Lys
734	4.43470	0.00504	Ha	727Lys
735	2.87564	0.00774	Hea	727Lys
736	1.67638	0.00852	Hbb	727Lys
737	1.36656	0.00648	Hba	727Lys
738	28.77083	0.09669	Cd	727Lys
739	0.98645	0.00613	Hga	727Lys

Number	Shift	SD	Assign Name	Residue
740	1.12396	0.00567	Hgb	727Lys
741	1.59573	0.00541	Hdb	727Lys
742	42.03510	0.03771	Ce	727Lys
743	1.59391	0.00704	Hda	727Lys
744	2.87593	0.00582	Heb	727Lys
745	8.34036	0.00920	H	728Thr
746	119.80928	0.03450	N	728Thr
747	61.40821	0.04774	Ca	728Thr
748	70.13314	0.05542	Cb	728Thr
749	173.03534	0.00841	C	728Thr
750	21.84853	0.03024	Cg2	728Thr
751	1.03989	0.00541	Hg2*	728Thr
752	5.26167	0.00724	Ha	728Thr
753	3.84826	0.00390	Hb	728Thr
754	9.23645	0.00848	H	729Tyr
755	127.20205	0.02256	N	729Tyr
756	56.44687	0.09072	Ca	729Tyr
757	41.59646	0.05059	Cb	729Tyr
758	175.06743	0.00487	C	729Tyr
759	4.83400	0.00524	Ha	729Tyr
760	3.17325	0.00577	Hbb	729Tyr
761	2.62043	0.00445	Hba	729Tyr
762	6.85247	0.00626	He*	729Tyr
763	7.27247	0.00889	Hd*	729Tyr
764	134.67803	0	Cd*	729Tyr
765	118.14031	0	Ce*	729Tyr
766	8.62222	0.00971	H	730Thr
767	116.26265	0.01340	N	730Thr
768	61.30505	0.03163	Ca	730Thr
769	71.06038	0.07727	Cb	730Thr
770	173.82426	0.00588	C	730Thr
771	21.03381	0.03542	Cg2	730Thr
772	1.04607	0.00476	Hg2*	730Thr
773	5.44375	0.00796	Ha	730Thr
774	3.91637	0.00167	Hb	730Thr
775	9.04489	0.00913	H	731Glu
776	125.59238	0.01757	N	731Glu
777	54.48766	0.08963	Ca	731Glu
778	33.19122	0.06085	Cb	731Glu
779	174.75107	0.00211	C	731Glu
780	34.90968	0.04837	Cg	731Glu
781	4.71397	0.00417	Ha	731Glu

Number	Shift	SD	Assign Name	Residue
782	1.56034	0.00265	Hba	731Glu
783	2.22629	0.00978	Hgb	731Glu
784	2.13973	0.00777	Hga	731Glu
785	2.02868	0.00586	Hbb	731Glu
786	8.70719	0.00870	H	732Lys
787	124.57960	0.01194	N	732Lys
788	55.50923	0.08485	Ca	732Lys
789	33.55290	0.04978	Cb	732Lys
790	176.19420	0.02156	C	732Lys
791	25.05923	0.09467	Cg	732Lys
792	41.69436	0.08010	Ce	732Lys
793	29.01491	0.05907	Cd	732Lys
794	5.17700	0.00625	Ha	732Lys
795	2.89006	0.00632	Hea	732Lys
796	1.68482	0.00981	Hba	732Lys
797	1.76439	0.00979	Hbb	732Lys
798	1.47659	0.00591	Hgb	732Lys
799	1.24616	0.00786	Hga	732Lys
800	1.65572	0.00955	Hda	732Lys
801	1.68471	0	Hdb	732Lys
802	2.89198	0.00316	Heb	732Lys
803	9.11982	0.01919	H	733Val
804	123.42393	0.08238	N	733Val
805	35.32762	0.05537	Cb	733Val
806	173.54989	0.01199	C	733Val
807	60.55927	0.04261	Ca	733Val
808	1.91190	0.00561	Hb	733Val
809	4.53072	0.00663	Ha	733Val
810	21.95437	0.07373	Cgb	733Val
811	0.78048	0.00268	Hga*	733Val
812	0.80643	6.56863e-04	Hgb*	733Val
813	20.98310	0.03916	Cga	733Val
814	8.46059	0.00894	H	734Ile
815	125.31105	0.03348	N	734Ile
816	60.79610	0.04234	Ca	734Ile
817	39.87838	0.10726	Cb	734Ile
818	174.92111	0.01834	C	734Ile
819	28.42270	0.07331	Cg1	734Ile
820	17.60563	0.01756	Cg2	734Ile
821	13.74112	0.02861	Cd1	734Ile
822	4.67375	0.00307	Ha	734Ile
823	1.78038	0.00495	Hb	734Ile

Number	Shift	SD	Assign Name	Residue
824	1.60228	0.00469	Hg1b	734Ile
825	1.12255	0.00310	Hg1a	734Ile
826	0.83753	0.00171	Hg2*	734Ile
827	0.91814	0.00454	Hd1*	734Ile
828	9.08129	0.00946	H	735Val
829	129.17787	0.02293	N	735Val
830	32.73268	0.02348	Cb	735Val
831	62.43732	0.09061	Ca	735Val
832	174.87364	0.02687	C	735Val
833	4.16852	0.00607	Ha	735Val
834	2.44900	0.00622	Hb	735Val
835	0.85100	0.00700	Hga*	735Val
836	0.91221	0.00391	Hgb*	735Val
837	21.87564	0.02733	Cgb	735Val
838	21.44569	0.00596	Cga	735Val
839	8.03985	0.00836	H	736Lys
840	132.67456	0.02120	N	736Lys
841	34.41198	0.04863	Cb	736Lys
842	57.25529	0.07769	Ca	736Lys
843	170.43650	0	C	736Lys
844	4.27379	0.00556	Ha	736Lys
845	41.94342	0.06415	Ce	736Lys
846	24.50601	0.09587	Cg	736Lys
847	29.12689	0.07635	Cd	736Lys
848	1.74425	0.00441	Hbb	736Lys
849	2.98613	0.00639	Heb	736Lys
850	1.65391	0.00912	Hdb	736Lys
851	1.36823	0.01307	Hgb	736Lys
852	1.33904	0.00688	Hga	736Lys
853	1.74271	0.00508	Hba	736Lys
854	1.65383	0.01068	Hda	736Lys
855	2.98603	0.01035	Hea	736Lys